FACULTY OF PSYCHOLOGY AND EDUCATIONAL SCIENCES

Can machine learning capture differences in EEG of infants at elevated likelihood and typical likelihood of Autism?

Word count: 18,399

Sam Boeve
01709051

Promotor: Prof. dr. Herbert Roeyers
Supervisor: Ijlal Haider

A dissertation submitted to Ghent University in partial fulfilment of the requirements for the degree of Master of Science in Psychology

Academic year: 2021 – 2022

GHENT UNIVERSITY

Corona Preamble

The Covid-19 pandemic had an impact on the completion of this dissertation. The entirety of the EEG data was collected beforehand by a third party. There were, however, many missing scores on the screenings of the infants at 24 months old since the Covid-19 measures prevented the administration of those instruments. This has led to a smaller sample size that could be used for the present analyses. The pandemic also had an influence on the efficiency in communication between myself and my supervisor, Ijlal Haider. Despite these obstacles I strived to make this thesis a success.

Abstract

Autism spectrum disorder (ASD) is a group of neurodevelopmental disorders marked by impairments in the social domain and in behavioural flexibility. The diagnosis of ASD is based on clinical judgement. Due to the nature of the impairments (social, communicative, behavioural) and the heterogeneity thereof, an early diagnosis and start of treatment remains cumbersome. Past studies showed that the inner dynamics of the brain could be among the earliest biomarkers of autism with signs emerging within the first year of life. In the present study the potential of machine learning algorithms for the classification of EEG derived features at 10 or 14 months old (both simple and complex) as belonging to infants with a positive or a negative screening for autism at 24 months old, is evaluated. For this study, EEG data was collected from infants at elevated risk of autism (N = 64) at 10 and 14 months old. Analyses revealed that support vector machines were not able to predict the result of the screening conducted at 24 months based on features derived from EEG recordings at both 10 and 14 months of age. The sensitivity and specificity of the models were insufficient. A large scale comparison of the EEG features from the group with a positive screening at 24 months showed no significant differences to those of the group with a negative screening. The implications of these results and future directions for research on endophenotypes of autism are discussed.

Keywords: autism, EEG, support vector machines, prediction, early diagnosis

Samenvatting

Autismespectrumstoornis (ASS) is een overkoepelde term voor een verzameling van ontwikkelingsstoornissen die gekenmerkt worden door beperkingen in het sociaal functioneren en in de gedragspatronen. De diagnose van ASS wordt gesteld op basis van een klinisch oordeel. Vanwege de aard van de beperkingen (sociaal en gedragsmatig) en de heterogeniteit daarvan, blijft een vroege diagnose en start van behandeling moeilijk. Voorgaand onderzoek toonde aan dat een afwijkende hersendynamica een van de vroegste tekenen kan zijn van ASS, waarbij deze al op te sporen zijn in het eerste levensjaar. De huidige studie onderzoekt het potentieel van *machine learning* algoritmes voor de classificatie van EEG waarden van 10 maand oude kinderen (gaande van simpele tot complexe waarden) als behorende tot een kind met positieve of met een negatieve screening voor autisme op een leeftijd van 24 maanden. In het kader van dit onderzoek werden EEG opnames afgenomen bij 64 kinderen met een verhoogde kans op autisme toen deze 10 en 14 maanden oud waren. Analyses toonden aan dat *support vector machines* niet in staat waren om het klinische oordeel, dat gesteld werd op een leeftijd van 24 maanden, te voorspellen op basis van de EEG waarden. De sensitiviteit en specificiteit van de modellen was ontoereikend. Een grootschalige vergelijking van de EEG waarden van de groep met een positieve screening op 24 maanden en deze met een negatieve screening toonde geen significante verschillen aan. De implicaties van deze resultaten en richtlijnen voor toekomstig onderzoek naar de neurobiologische basis van autisme worden besproken.

Kernwoorden: Autisme Spectrum Stoornis, EEG, support vector machines, voorspelling, vroege diagnose

Prologue

Lo and behold, the thesis 'Can machine learning capture differences in EEG of infants of elevated likelihood or typical likelihood of autism?'. This thesis has been established through the cooperation of many actors. It is part of the 'Tracking Infants at Risk for Autism' or TIARA project. The project, a large scale collaboration between Ghent University and KU Leuven, aims to understand why some children develop autism while others do not. Finding early markers of this disorder opens many opportunities for early assistance and counselling and thus improving the quality of life of those involved. The researchers who administered all the EEG recording deserve great praise as this is an immense undertaking, with many additional difficulties during times of a pandemic. Another actor who deserves praise for his role in this project is the promotor of this thesis, Prof. dr. Roeyers. I want to thank him for his calm and thoughtful assistance when the timely completion of this project seemed unfeasible due to problems acquiring the data. Lastly, but most importantly, I want to thank my supervisor, Ijlal Haider. Ijlal's help in extracting the EEG features and his knowledge about signal processing was indispensable. He deserves credit for his perseverance in getting the high performance computing infrastructure to work with us, his swift learning of Linux and his ever optimistic attitude. Thank you. Furthermore, I need to apologize to all friends and family members whom I have had read this thesis repeatedly, your struggles are over.

As an experimental psychologist, my deep interest for the inner workings of the brain and my love for data analysis pushed me towards this thesis. Both these initial motivators are deeply fulfilled. Looking back on this thesis, I can honestly say I have learned more than I could have hoped.

Before you lies a thesis of which I'm very proud. I hope it may inform you and provide a useful contribution to the research on autism. I wish you great reading pleasure.

Sam Boeve,

Ghent, May 24th 2022

Autism Spectrum Disorder

Diagnosis of Autism: Current Affairs

Behavioural Symptoms

Autism spectrum disorder (ASD) is a highly debated and a widely studied topic. This comes as no surprise since this disorder consists of a vast amount of developmental trajectories and behavioural phenotypes. The modern concept of autism was first coined by the Austrian – American psychiatrist Leo Kanner. In his 1943 paper, "Autistic Disturbances of Affective Contact", he describes the cases of 11 children who he deemed to have no predisposition to be social. In his analysis of this disorder he also noted the trouble the patients experienced in non-social aspects of life like their 'insistence on sameness' and labelled some of the repetitive patterns of behaviour, e.g. body rocking, as attempts of the child to resist change. In this paper, he also noted the problems many of the subjects experienced in speech or language related processes such as echolalia. However, he did not identify these problems as an essential part of autism. Following this initial report of Kanner (1943), a gradual increase of research into the diagnosis of autism and neurobiological and developmental mechanisms underlying this disorder took place. In these years a debate started on the definition and criteria of autism, together with a consensus for the need of more formal diagnostic guidelines. An overview of the changes in criteria and nomenclature through the years is beyond the scope of this thesis, but we refer the interested reader to Volkmar and McPartland (2014) for an excellent review.

The latest changes with regard to the criteria and diagnostic labels were made in the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association [APA], 2013). In this edition autism spectrum disorder is the broad category which harbours the previous distinct developmental disorders such as autistic disorder, Asperger's disorder, pervasive developmental disorder – not otherwise specified (PDD – NOS) and childhood disintegrative disorder (CDD). Two domains of impairment are identified in DSM-5, namely social/communicative difficulties and restricted/repetitive behaviours. The impairment in social communication and social interaction consists of three clusters of symptoms. This domain is monothetic, meaning that a person must display symptoms from all three clusters to meet the criteria of ASD.

1. Deficit in social-emotional reciprocity. Symptoms belonging to this cluster are usually expressed as a lack of social initiation, unusual social initiations (e.g. intrusive touching), failure to engage in a normal conversation, etc.
2. Deficits in nonverbal communicative behaviours used for social interactions. Typical symptoms from this cluster are the inability of social eye-contact, abnormal volume, pitch, intonation or prosody in speech.
3. Deficits in building and maintaining relationships. This deficit manifests itself in the lack of theory of mind, the lack of interest in others, etc.

This social component needs to be accompanied by some form of restricted, repetitive patterns of behaviour, interests or activities (e.g. stereotyped or repetitive speech). This domain is characterized by four clusters. Contrary to the social domain, this domain is polythetic, meaning that a person must show symptoms from two out of the four clusters in order to meet the criteria.

1. Stereotyped or repetitive speech (e.g. echolalia or overly formal language), motor movements (e.g. swaying) or use of objects (e.g. non-functional play with objects).
2. Excessive adherence to routines or an aversion for change (e.g. overreaction to trivial changes).
3. Highly specific interests that are overly intense (e.g. preoccupation with number or letters).
4. Hypo – or hyper-reactivity to sensory aspects of the environment (e.g. inappropriate smelling of objects).

On top of the deficiencies in both the aforementioned areas, the symptoms must be present during early childhood. Early childhood is defined flexibly as eight years of age or younger. However DSM-5 notes that the symptoms may not fully manifest themselves until later when social demands increase. Lastly, these symptoms should interfere significantly with everyday life to meet the diagnostic criteria of the DSM-5. Based on the severity of the impairment in social and behavioural areas and the need for support in everyday functioning, three levels of severity are specified with the third level indicating that substantial support is needed.

A study by McPartland et al. (2012) examined the effect of these new criteria on the sensitivity and specificity of the diagnosis of autism. They re-evaluated a sample of more than 900 participants that was used in a study using DSM-IV criteria (APA, 1994). The sensitivity, the ability of the new criteria to correctly identify those with autism, identified using the previous criteria (i.e. true positive rate) was 60.6%. It was found that the milder forms of autism, subgroups like Asperger's disorder or PDD-NOS, were less likely to be labelled as such. The specificity, however, the ability of the new criteria to exclude those without autism, was high, 94.9%. Across studies it was shown that the new criteria tend to include only the individuals with more severe impairments (Worley & Matson, 2012; Matson et al., 2012a). Thus there appears to be an increased specificity at the cost of less sensitivity.

The diagnosis of autism is based on clinical judgement. There is however a need for diagnostic tools in order to structure and objectify this judgement. A great deal of effort went into the research and development of a reliable tool to help a clinician make a diagnosis. At first, this effort led to a whole series of questionnaires, observation methods, screening instruments and scales.

Diagnostic Instruments

One of the first attempts to develop a reliable scale to objectively identify cases of autism was made by Rimland (1964). The result of his attempt was the Form E-1, a diagnostic scale containing a range of meticulously selected symptoms based on the concept of autism as described by Kanner (1943). This made it possible for clinicians to more objectively diagnose a patient with autism and it allowed researchers to compare cases. Form E-1 was originally designed to differentiate between children with autism and children with schizophrenia. Later, the measure was revised to the Form E-2 to tell children with and without autism apart. This form contained 79 items related to the behaviour of the child and relies on retrospective information from the parents (Butler & Lord, 2013).

A more recent and highly influential diagnostic tool was developed by Lord et al. (1989). They designed and evaluated the Autism Diagnostic Observation Schedule (ADOS). ADOS is a standardized protocol for observation of social and communicative behaviour and play/imagination normally associated with autism. This protocol consists of a series of structured and semi-structured tasks which induce the need of social interaction between the person under examination and the examiner. On average it takes 30 to 60 minutes to administer. In 2012, a revision of this tool was published, namely the ADOS-2 (Lord et al., 2012). ADOS-2 contained updated norms and the different modules were slightly altered to improve the accuracy and expanded to accommodate for a larger client population. It consists of five modules, which are selected based on the level of expressive language of the examinee and on his/her age. Module four is suited for verbally fluent adults or older adolescents. The toddler module, on the other hand, is designed for children aged 12-30 months and who are not able to form sentences. Based on the reported data in the original study of (Luyster et al., 2009), the toddler module had a sensitivity of .88 and a specificity of at least .90. The inter-rater reliability was assessed to be satisfactory.

The tasks are designed to structure the interaction between the examiner and examinee but should be administered dynamically so that the social interaction with the examinee is natural. Coding of the interactions should follow the examination, based on the notes of the examiner. Scoring is done with regards to the language and communication, reciprocal social interaction, play and imagination and stereotyped behaviours displayed by the subject. A code ranging from 0-2 indicates the degree of abnormality of displayed behaviour. For the first three modules (i.e. toddler module, 1, 2) a parent or caregiver must be present in the room during the observation. The ADOS-2 is a level C measure, meaning it can be administered and interpreted by professionals from medicine, psychology or a related field. Before conducting the observation the administrator must have completed the appropriate clinical training.

The strong points of ADOS-2 are the different modules based on language ability, the two broad domains of social affect and restricted behaviour which align nicely with the DSM-5 conceptualization.

This, in combination with the enhanced sensitivity and specificity, make this a valid diagnostic tool (McCrimmon & Rostad, 2013). The fact that it is widely used, makes it ideal for comparison and research. There is also a Dutch adaptation of the ADOS-2 available (de Bildt et al., 2013). This diagnostic instrument received the highest quality label by the *Kwaliteitscentrum voor Diagnostiek vzw* (2018), a Flemish non-profit organisation aimed at developing and implementing scientifically supported diagnostic instruments.

Some of the limitations of ADOS-2 are the standardization sample. This sample consists primarily of male, Caucasian Americans which limits the generalizability of the tool. Finally, the results of the ADOS are dependent on the clinical skills of the clinician administering the diagnostic tool. It is therefore of great importance that the clinician received sufficient training.

While these diagnostic tools are aimed at individuals showing some symptoms of ASD in order to establish a diagnosis, the present study involves young children potentially showing very little to no symptoms. Therefore, there is the need for reliable screeners to detect at-risk individuals.

Screening Instruments

The CHecklist for Autism in Toddlers (CHAT) was one of the first validated screenings available for autism. The original study (Baron-Cohen et al., 1992) aimed to establish if detection of high-risk infants was feasible for toddlers at 18 months. The CHAT was designed to be administered by healthcare professionals and consisted of several items aiming to assess each of six areas of development (e.g. social play, joint attention, etc.). The initial results showed great potential as the screening at 18 months of age showed a relatively high positive predictive value of the diagnosis at 2.5 years old. Later, the results of the follow-up, when the participating children were 6 years old, painted a different picture as only 38% of children with a diagnosis of ASD were identified by the CHAT (Baron-Cohen et al., 1996).

The limitation of the CHAT led to the development of the Modified CHecklist for Autism in Toddlers (M-CHAT; Robins et al., 1999). This checklist consists of 23 items and relies on the parents' report to evaluate the likelihood of autism in children aged 18-24 months. A major limitation of this type of screening is that it depends on the report of the parents, which is open to bias and exaggeration or minimization. Besides, since it is not standardized, it is prone to under-detection of children with ASD (Weitlauf et al., 2015). This screening instrument is provisionally recommended by the *Kwaliteitscentrum voor Diagnostiek vzw* (2018).

Out of the M-CHAT, the Quantitative CHecklist for Autism in Toddlers was born (Q-CHAT; Allison et al., 2008). The Q-CHAT was the next modification and changed the dichotomous responses (i.e. yes, no) on the items of the M-CHAT to ordinal responses, allowing for the indication of how much or how often the items manifested themselves in the child. This change to ordinal responses also

acknowledged that autistic traits lie on a dimension. The Q-CHAT was modified even further by Allison et al. (2012) as the 10 items which resulted in the highest test accuracy where selected to form a 'red flag' tool. Based on a sample of 126 toddlers and using a cut-point of 3 on the Q-CHAT-10, this red-flag version of the Q-CHAT had a sensitivity of .91, a specificity of .89 and a positive predictive value of .58 (Allison et al., 2012).

A screening tool for slightly older children is the Social Responsiveness Scale-2 (SRS-2; Constantino & Gruber, 2012). The age range for the SRS-2 is from 3 to 18 years, with separate scales for 3 year olds and for 4-18 year olds. The SRS-2 provides an indirect judgement of aspects related to ASD as perceived by the parents or caretakers. It consists of 65 questions related to interpersonal relationships, communication and stereotypical behaviour. Validation of this screening tool in a sample of school aged children showed a sensitivity value of .92 and a specificity of .92 meaning it does well in identifying at risk children (Bruni, 2014).

In summary, over the years it has been clear that designing a diagnostic tool that captures all forms of autism or a screener detecting at risk infants is challenging. Nevertheless, many good instruments are available. To form a diagnosis, the widely used and accepted ADOS-2 offers a strong and reliable framework to help clinicians. With regard to the screeners, the M-CHAT or Q-CHAT and the SRS-2 seem to be the most reliable options. There remains room for improvement in the research for an early detection tool. A major caveat is the absence of a behavioural marker before the first year of life. It is within this early stage of the development of ASD that the use of neuroimaging and related analysis techniques offers future avenues for research into the early prediction. The need for a reliable early detection method seems more pressing than ever, since some have argued that an autism epidemic is developing (Chiarotti & Venersosi, 2020). What are the facts behind these sensational claims?

Prevalence

Speaking of an epidemic of an disorder as heterogeneous as autism should be done carefully since it is very difficult to estimate the prevalence rates reliably in the first place. It should come as no surprise then that the prevalence rates have been a topic of much debate.

Williams et al. (2006) made a systematic review of prevalence studies on autism spectrum disorder based on the then used DSM-IV classification. The authors found that the overall prevalence across studies for typical autism was 7.1 per 10,000 (95% confidence interval (CI): 1.6 – 30.6) and of all ASD was 20.0 per 10,000 (95% CI: 4.9 – 82.1). It is worth noting that the variation in the studies estimating the prevalence of typical and all ASD was very large, further emphasising the heterogeneity of the disorder and the difficulties this poses for the diagnosis. For the entirety of the spectrum, the worldwide median is estimated to be around 62.0 per 10,000 while the European median is 61.9 per

10,000 (Elsabbagh et al., 2012). This assessment was updated in 2022 using the estimates of 71 studies (Zeidan et al., 2022). The result was a worldwide median prevalence of 100.0 per 10,000 (range: 1.09/10,000 to 436.0/10,000). The prevalence estimate was the same when online studies in Europe were included.

A review by Tsai (2014) dived into the prevalence rates of autistic disorder and autism spectrum disorder and investigated how the DSM-5 criteria could change the prevalence rates of ASD. While the prevalence of autistic disorder (i.e. narrow definition of autism, DSM-IV) is estimated in the range of 10 to 30 individuals per 10,000 people, the median prevalence of ASD, (i.e. encompassing autistic disorder, Asperger disorder and PDD-NOS, DSM-IV) is estimated to lay around 69.5 per 10,000 people. Because of the wide range of prevalence estimates of ASD, it is problematic to pick one particular prevalence. In this review, a sample of thirteen studies showed that 9% to 54% of individuals with a DSM-IV diagnosis of autistic disorder, Asperger disorder or PDD-NOS are not identified as having a DSM-5 ASD diagnosis. Because of this, a drop in prevalence is likely to be observed in the years of the transition and estimates of autism prevalence using the DSM-IV are not readily comparable to those using DSM-5.

Contrary to the idea that the use of the DSM-5 criteria would cause a drop in prevalence of ASD, in recent years prevalence of ASD seems on the rise. This has caused some researchers to speak of an autism epidemic. However, since ASD prevalence estimates show a very large variability, one should be careful when proclaiming the surge of an autism epidemic without looking into the possible reasons of the observed changes (Chiarotti & Venerosi, 2020). In Europe, the observed prevalences range from 42 per 10,000 people in Piemonte, Italy to 313 per 10,000 people in Iceland. In North-America this ranges from 100 to 185 per 10,000 people. Some factors influencing the variability of ASD prevalence, are the source of data used to estimate the prevalence, geographical area and additional environmental factors. The main sources of ASD prevalence are administrative data, ad hoc studies and surveys based on questionnaires. Parents and teachers reports seem to significantly overestimate the prevalence of ASD. With regard to the geographical data, it is noted that Europe shows significant lower prevalence of ASD than Australia and Asia. The geographical factor only accounts for about 50% of the variability, suggesting other factors are linked to the occurrence of autism as well. Current literature suggests several factors that could potentially affect brain development and differentiation during the perinatal period. The CHARGE (Childhood Autism Risks from Genetics and the Environment) study of Hertz-Picciotto et al. (2006) obtained evidence for folate prenatal intake, maternal fever, pesticide exposure and air pollution to be associated with an elevated likelihood of ASD. In addition to these environmental factors, the biggest risk factor seems to be the occurrence of an older sibling diagnosed with autism, preterm birth or feeding problems .

Infants at Elevated Likelihood

Feeding problems. Children with developmental disabilities are at elevated risk of developing feeding problems (Babbitt et al., 1994). For ASD specifically, it has been estimated that up to 89% of children with ASD show some sort of feeding concern. This ranges from strong preferences for certain foods (e.g. by type, texture, colouring, etc.), to consuming less food and a less diverse range of foods compared to peers (Ahearn et al., 2001; Schreck et al., 2004). Studies investigating a potential gastrointestinal etiology found no evidence of such a causality. This led to the hypothesis that the deviant feeding patterns among children with ASD are a consequence of the restricted interests and behavioural rigidity (Ledford and Gast, 2006). Be that as it may, there are reports of feeding problems starting in the first year of life of infants later diagnosed with ASD, well before any of the aforementioned symptoms manifests itself (Keen, 2008).

Recurrence in siblings. It is known that infants with an older sibling diagnosed with autism are at elevated likelihood of developing autism themselves. While the genetic basis of autism is immensely complex and heterogeneous, identifying such a genetic etiology can provide the clinicians and the family of the affected individuals with important information about the prognosis and recurrence risk (Hyman et al., 2020). Etiological investigations include a thorough medical and developmental family history, a physical examination of the child with, among others, assessment of growth (e.g. head circumference) relative to typical curves. Some clinical genetic syndromes such as fragile X syndrome or Rett syndrome in girls also meet the criteria for ASD. In their paper, Hyman et al. (2020) list the clinical syndromes and the most commonly identified DNA duplications or deletions that are associated with ASD. Although this kind of genetic information can be of high value for the parents, it is important to note that the diagnosis of ASD is made on the basis of clinical symptoms. When a specific genetic cause of the disorder is identified, the parents can be informed about the likelihood of recurrence in subsequent children.

In the absence of a clear origin or when this kind of investigation is deemed unnecessary, the likelihood of recurrence can be estimated. The likelihood of ASD in a child with one older sibling diagnosed with ASD is estimated at 18.7% (95% CI: 13.3–25.5). Couples with two or more children with a diagnosis of ASD have a chance of 32.2% (95% CI: 21.8–44.7) of having a subsequent child with ASD (Ozonoff et al., 2011). The recurrence rates in monozygotic and dizygotic twins are generally much higher. A meta-analysis by Tick et al. (2016) revealed co-occurrence rates as high as 53% to 67% in dizygotic twins and up to 98% in monozygotic twins. Siblings of a child with ASD are also more likely to develop symptoms related to ASD that do not meet the threshold for a clinical evaluation of ASD. The mapping of all the genetic effects that play a role in the expression of ASD is far from complete, yet it is clear that there is a strong genetic contribution to the ASD likelihood.

Preterm infants. Not only is there an increased risk of autism in the group of children with a sibling diagnosed with autism, also infants born preterm have an elevated likelihood of developing autism. An infant is said to be preterm when it is born before 37 weeks gestation. A meta-analysis by Agrawal et al. (2018) estimates the prevalence of ASD in preterm infants at 7% (95% CI: 4% - 9%). An estimate considerably higher than in the general population, where an overall prevalence rate of 1% has been reported (Zeidan et al., 2022). A recent study provides a tentative mechanism for the increased prevalence rate in preterm children (Bokobza et al., 2019). One important factor, responsible for many of the preterm births, is maternal-foetal inflammation (Goldenberg et al., 2008; Yeh et al., 2017). Subsequently, this can lead to foetal/neonatal inflammation which affects the normal neurodevelopment of the child (Levinton et al., 2018). This is due to the microglia, which are essential for normal synaptogenesis, being shifted away from their normal function of synaptogenesis towards a pro-inflammatory phenotype of the microglia due to the inflammation (Tay et al., 2017). This shift away from normal synaptogenesis potentially leads to deficits in neural connectivity. For example, it has been demonstrated that the preterm infants show a decreased connectivity between cortex and thalamus compared to term infants (Ball et al., 2013). More evidence regarding the deficits of neural connectivity in ASD are reviewed in the section on functional connectivity (p. 20).

The prevalence of ASD in the population is relatively low and the first behavioural signs of autism only manifest themselves at a later age, making research on the neurodevelopmental trajectory of ASD resource intensive. Studying the population of infants with feeding problems, a diagnosed sibling or preterm infants offers the possibility of studying a sample with many more participants eventually receiving the diagnosis of ASD.

Developmental trajectory

Studies investigating the developmental trajectory of autism in infants at an elevated likelihood have run into much of the same obstacles as the prevalence studies. To study the evolution over time we need a measure suited for participants of different ages and symptom severity and one that is preferably comparable across subjects and time.

This issue involves reaching a consensus on a measure suitable for longitudinal comparisons. As was mentioned in the section on the diagnosis of autism, the Autism Diagnostic Observation Schedule or ADOS is widely used and commonly viewed as the golden standard in the assessment of autism. For many years, the ADOS raw totals have been collected and reported. However, for the purpose of comparing treatment outcome and longitudinal change, a related measure, the ADOS calibrated severity score (ADOS-CSS; Gotham et al. (2009)) is more appropriate since this measure aims to minimize influences of chronological age, language and overall cognitive ability.

Bieleninik et al. (2017) published a meta-analysis on the temporal stability of autism spectrum diagnosis and severity. Combining the results of 40 studies, they reported a significant overall improvement in the severity of the symptoms as measured by the ADOS total scores. Yet, no significant change was observed when they used the ADOS-CSS as a measure of severity. The authors interpreted this disparity as an indication that using the ADOS total scores may result in artificial improvements due to changes in the ADOS modules and overall development of the participant.

Overall these results indicate the strong temporal stability of ASD even when individual trajectories do show great variability. This study also emphasized the use of ADOS-CSS. This score is one of the more robust tools to be used in the assessment of the evolution of autism and as such it consists vital tool for the evaluation of potential and established interventions.

Intervention

Types of Interventions. Designing an effective treatment has proven difficult in the past. Autism is a pervasive and rigid developmental disorder marked by a stability in severity over time. Yet, there is some convergence in the literature on autism treatment between the behavioural and developmental approaches (Narzisi et al., 2014).

The former focuses on learning adaptive behaviours and is often based on the principals of reinforcement. This approach is marked by a more formal context in which a therapist structures the interaction. Interventions of this type are Applied Behavioural Analysis (ABA; Lovaas (1987)) also known as Early Intensive Behavioural Intervention (EIBI), Pivotal Response Training (PRT; Koegel et al. (1998)), Picture Exchange Communication System (PECS; Bondy & Frost (2001)). The latter, the developmental method, focuses more on the individual trajectory of each child and the child's initiative. This approach emphasises the merits of a more naturalistic context with a substantial role for the parents. Interventions like the Early Social Interaction (Wetherby & Woods, 2006) and the SCERTS program (Prizant et al., 2003) can be classified as developmental in nature. In order to evaluate these interventions, there is a need of solid outcome measures.

Outcome. There are a number of candidate outcome measures. A crucial outcome measure was mentioned above, the ADOS calibrated severity scores. This measure is relatively insensitive to age, language development or overall cognitive ability.

A second measure that is reported regularly, is the intelligence quotient (i.e. IQ). However, since IQ is prone to increase over time within an individual without having a positive effect on her/his social functioning, this measure alone is viewed as insufficient to determine outcome.

An important addition in the evaluation of the treatment is the inclusion of outcome measures related to overall well-being and family quality of life. A meta-analysis revealed that people with autism experience a lower quality of life compared to people without autism (van Heijst & Geurts, 2015). This

indicates the potential of 'quality of life' as a focus for interventions and the need for a better understanding of this construct and the factors contributing to it (Renty & Roeyers, 2006).

Effectiveness of Early Interventions. Based on these outcome measures, many studies have looked into the effects of treatment. These studies on the effectiveness of early autism intervention are without doubt a laudable undertaking, yet they are not without controversy. Four main issues complicate the usefulness and the interpretation of the studies investigating interventions of autism (Reichow, 2012).

The first is the lack of consistent use of an outcome rating. As such, this poses a major challenge in generalizing the findings (Magiati & Howlin, 2001). Second, many studies investigating the outcome of autism treatment lack in some methodological aspects. Different studies use different research designs with a wide range of comparison groups. Third, besides the difficulties posed by this variability, many studies fail to report the criteria for selection and evaluation of patients and the allocation of the participants to the different groups. This issue also encompasses the lack of a standardized autism diagnostic instrument. How can one study the effectiveness of an intervention when merely identifying the subjects at a young enough age is problematic? Lastly, the well-known publication bias adds to the issues.

A review looking into the effectiveness of early interventions painted a rather grim picture (Warren et al., 2011). This review included 34 studies. The authors concluded that the overall strength of evidence ranged from insufficient to low. There is, however, reason to belief that the lack of solid evidence of an effective treatment in these studies results from other issues than the interventions themselves. A total of 23 studies were rated as being of poor quality according to certain predetermined characteristics. A similar result was obtained in a recent individual participant data meta-analysis (Rodgers et al. 2021). There was some evidence that early intervention may lead to larger improvements in child cognitive ability and behaviour. However, as in most studies, there was a considerable variance among participants and all studies were identified as having a risk of bias.

Although the research on the effectiveness of the interventions is plagued by methodological shortcomings, there are some sound motivations to develop tools that help in the early detection of autism. Not only will an improved early diagnostic sensitivity improve the outcome for the individuals involved, it will also improve the studies looking into the effectiveness of the interventions and the conclusions that can be drawn from them. The major rationale for the development of new detection tools, is the notion that autism is characterized by a prominent deficit in social interaction together with a tendency of focussed and intense interests. This causes the child to miss important learning cues from parents, siblings or peers. Reducing the pervasiveness of the symptoms in the developmentally crucial early years should result in improved outcomes (Camarata, 2014).

Neuroimaging and Autism Research

The impetus to start treatment as soon as possible accelerated the quest to predict/diagnose autism as early as possible. This has spurred an interest in endophenotypes of autism. Endophenotypes are intermediators between genotype and phenotype. “An endophenotype is a quantitative biological trait that is reliable in reflecting the function of a discrete biological system and is reasonably heritable, and as such is more closely related to the root cause of the disease than the broad clinical phenotype” (Wong et al., 2011, p. 338). Endophenotypes can range from biochemical to neurophysiological to behavioural traits.

Techniques such as electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRS) are becoming increasingly popular in infant studies on the development of autism for the identification of such biomarkers. Although the majority of past studies focussed on school-aged children, a focus shift has been seen towards studies with toddlers and infants. This has led to the accumulation of evidence that neuroimaging techniques are crucial in the early detection of autism. In our discussion on the potential endophenotypes of autism, we will focus on functional brain alterations typical for ASD as measured using EEG during resting state activity

Infant Studies and Neuroimaging

EEG forms an excellent measure to study brain function in developmental disorders such as ASD. Magnetic resonance imaging (MRI) and magnetoencephalography (MEG) apply large machinery, combined with relatively loud noises that can be frightening for young participants to be in for an extended period of time. Second, compared to other brain imaging techniques, EEG has a relatively high tolerance for movement, since the cap is fixed on the head of the participant. Third, as EEG is very cheap to use, it is widely available with a large body of research supporting inference of the resting state recording. Fourth, as it is totally non-invasive, it permits researchers to collect longitudinal data from a single participant. A final advantage of using EEG combined with resting state recording is that it is not necessary for young participants to respond to a stimulus. This is particularly useful because many of the participants are too young to be able to perform a task or their cognitive or physical impairments could prevent them from adhering to a certain taskset. Moreover, task related differences in evoked potentials are hard or even impossible to interpret correctly without a thorough understanding of the functional impairments of participants with ASD. Adding upon this issue is the observation that the brain operates as a whole, with external information interacting rather than shaping the operation of the brain (Raichle & Snyder, 2007).

Combined, these features make EEG the most appropriate neuroimaging modality for research looking into possible endophenotypes of autism. The identification of such an endophenotype and a reliable method of measuring this endophenotype could greatly improve the youngest age at which one

can pinpoint an infant with an elevated likelihood of developing autism. Not only could insights in the neurological basis of the disorder improve the age at which an intervention can take place, these insights could also help in developing new and potentially more effective interventions.

Potential Biomarkers Derived from EEG

Oscillatory Power. A first potential biomarker is the oscillatory power of the EEG recording. The oscillatory activity of groups of neurons results from the interneuron feedback connections and leads to synchronized firing of the neurons. One of the most common ways to describe resting state EEG is to break it down in distinct frequency bands at which groups of neurons are oscillating. These bands range from 0.3 Hz to over 100 Hz. While the exact labels of the bins differ, a common way to categorize the frequency bands is delta (1 - 4Hz), theta (4 - 8Hz), alpha (8 - 16Hz), beta (16 - 32Hz), gamma (32 - 64Hz), high gamma (64 - 128Hz). There exists a large literature on the functions associated with the different frequency bands. Alpha waves are present in relaxed individuals and are often linked to cognitive inhibition. Beta waves are linked to alertness and so forth (Cohen, 2017). In measuring the band power, a researcher can report the relative power or the absolute power. Relative power of a frequency band is the amount of activity in that band relative to the amount of activity in all the other bands. Absolute power, the activity in one frequency band independent of the activity of other bands, is actually preferable to characterize the alterations occurring in the brain of patients suffering from ASD. This is because alterations in multiple bands may confound inferences made from the relative power measures. One measure that is used to quantify the absolute power in a frequency band, is the root mean square (RMS) value (Cornew et al., 2012).

A review by Wang et al. (2013) combines the results of a large collection of studies on the resting state activity of ASD. Despite the differences between studies, a fairly consistent pattern emerged on the neurological abnormalities associated with ASD. There appears to be excessive power at low (i.e. delta, theta) and high frequency bands (i.e. beta, gamma) together with a reduced power in the middle ranges of frequency (i.e. alpha). Besides the consistent finding of the U shaped abnormality on the frequency spectrum, the locus of this abnormality appears to be diffuse and widespread, involving regions as the dorsal midline, parietal, right temporal and frontal areas for the delta power. Theta power enhancement is primarily observed in the frontal and right posterior cortex. For the high frequency oscillation bands, the most significant alterations have been observed in the occipital and parietal regions. Across studies large differences in the EEG power in participants with ASD have been observed. This discrepancy, together with the observation that spectral power is apt to change over time in developing individuals, has led Tierney et al. (2012) to investigate the developmental trajectory of resting state EEG power. While all individuals at an elevated likelihood of ASD (i.e. siblings with ASD, preterm infants, etc.) showed lower power in all frequency bands, these differences disappeared by 24

months of age for the delta, theta and beta power. Only alpha and gamma frequency power remained different after this developmental window. This study points to the importance of age in the study of spectral power, since group differences across different ages might be misleading.

Other authors have proposed putative effects of the alterations in these frequency bands. Cornew et al. (2012) bring to attention that alpha activity is inversely related to perception and action. As such, it might be that the observed reduction in alpha power makes ASD patients less prepared to process sensory input. This is further corroborated by the finding that ASD is associated with a delayed M100 evoked response (Roberts et al., 2010). Future work investigating the functional correlates of the modulations in oscillatory power will help elucidate the mechanisms of dysfunction in ASD by shedding light on the cognitive and perceptual impairments underlying the symptoms.

Functional Connectivity. The second EEG derived property of brain functioning and a potential biomarker is functional connectivity. There are multiple measures of connectivity and the term is used interchangeably. While there are three general concepts of connectivity (Horowitz, 2003), namely structural, effective and functional connectivity, we will only discuss functional connectivity. Functional connectivity refers to the temporal coherence or correlation of activity of neurons or groups of neurons. EEG has properties which make it an excellent method for studying functional connectivity. It has a direct link to neuronal activity and combined with its high temporal resolution makes it a suitable neuroimaging technique in uncovering the role of dynamic activation and deactivation in ASD. We note, however, that the correlation between signals in different regions picked up by EEG, are not accounting for a potential shared connection to a third region which is responsible for the correlation (e.g. thalamo-cortical pathways).

Autism is often thought to have its origin in a long-range underconnectivity and a local over-connectivity (Vissers et al., 2012). A systematic review of studies, delving into the characteristics of functional connectivity in the brain of participants with ASD, supports a general trend of long range under-connectivity (O'Reilly et al., 2017). Regions most often associated with long range under-connectivity are the prefrontal cortex and the posterior cingulate cortex. Some other regions, which showed varying results with primarily long range under-connectivity, were the precuneus, anterior cingulate cortex, superior temporal gyrus, posterior superior temporal sulcus, anterior insula and the parietal lobe. The results with regard to local connectivity are less coherent with some studies showing under-connectivity, some show over-connectivity and some even report mixed results. These differences could be the result of condition specific effects. Indeed, a study by Orekhova et al. (2014) showed an elevated phase-lagged alpha-range connectivity at short and long distances in infants who later developed autism compared to typical likelihood controls and elevated likelihood individuals who did not develop autism. Studying local connectivity using EEG or MEG is more problematic since

comparison of nearby electrodes is confounded by volume conduction. Moreover, the cases in which a local over-connectivity is observed could be confounded by a reduction in the long range connectivity, making the local connectivity more observable without a true increase in local connectivity.

The information on the connectivity can be supplemented using specific frequency bands. Overall evidence indicates that the lower frequency bands are characterised by an under-connectivity, while the higher frequency bands are linked to a potential over-connectivity (O'Reilly et al., 2017). This general under-connectivity in the lower frequency bands is also linked to the number of autistic traits in neurotypical subjects (Barttfeld et al., 2013).

The review concludes by mentioning that ASD is characterized by a functional connectivity that, in general is more randomly organised compared to normal functioning subjects (O'Reilly et al., 2017). There is a substantial amount of evidence for long range under-connectivity. The case for local over-connectivity is less clear cut with probable condition specific effects. Across studies, generalisation is problematic due to sample characteristics (e.g. age, diagnosis, etc), choice of frequency band, reference electrode and choice of experimental paradigm (e.g. resting state, photic driving, visual stimulation, etc). None of the reported studies utilised machine learning to classify the connectivity patterns as belonging to infants at an elevated likelihood or typical likelihood of ASD.

Hemispheric Asymmetry. A third potential endophenotype is the altered hemispheric asymmetry of the brain in patients with ASD. A review on the abnormalities of resting state EEG (Wang et al., 2013) in individuals with autism reported a wide range of studies in which an increased power in the left hemisphere was observed for all frequency bands. This increased power spread throughout the cortex and covered large parts of the temporal, parietal and frontal regions. We note that the report of these abnormalities has been inconsistent with studies reporting reduced power in the left hemisphere (Dawson et al., 1995) and reduced power in the right hemisphere (Lazarev et al., 2009) in children with autism.

The increased power in the left hemisphere at rest in autism may contribute to the deficits in functions associated with this region, such as language. An increase in power at rest may reduce the signal-to-noise ratio during active tasks.

Nonlinear Measures of the EEG Signal. While the aforementioned analysis techniques and measures are associated with ASD, prediction based on these measures lacks accuracy (Bosl et al. 2018). This could be due to the fact that all these analysis techniques are linear and the brain functions in a nonlinear manner. Neurons are organised as a scale free network, meaning that the network characteristics are independent of the underlying number of neurons. A network is scale free when the distribution of the number of edges (i.e. the connections) per node follows a so called power law. The

consequence of this power law distribution of the edges is that you get a small number of highly connected nodes, the hubs. Neurons are thus organised as a scale free network meaning that the EEG signals carry nonlinear complex system information. The neurons act as many interacting nonlinear oscillators, which display chaotic and transiently synchronized activity. Understanding these dynamics is crucial to understanding the emergent properties of these dynamics such as behaviour and eventually to understand dysfunctions such as autism. Linear analysis techniques fail to detect these nonlinear properties of the neuronal information and thus miss out on a lot of information (Bosl et al., 2011). The potential for EEG to be used as a method to uncover and evaluate biomarkers of autism, continues to improve as new methods for analysing complex dynamic signals are being developed and integrated into research.

The first nonlinear features covered here are the values derived from the Recurrence Quantitative Analysis (RQA). RQA is a way to analyse and represent dynamical systems. It quantifies the number and the duration of recurrences of such a system (Marwan, 2012). The observation that certain states of a system will recur, meaning it will revisit that exact state up to an error, originated in astrophysics but has since then been used to study the properties of many different systems from financial markets to brain systems. These recurrences happen within the phase space of the dynamical system. A phase space represents all the possible states of a system, every parameter of this system (e.g. pressure, humidity, etc for physical systems) is represented on an axis in multidimensional phase space. In most cases, phase space has a dimension too large to depict it (Marwan et al., 2007). Recurrence plots are a way to visualize this high dimensional phase space. A recurrence plot visualizes the moments in time when the system visits through the same area in phase space. It is actually the visualisation of a square matrix where the matrix elements correspond to those times at which a state of a dynamical system repeats itself, in other words when it recurs (Marwan, 2012). The recurrence of a state is marked with ones (i.e. equality of the states up to an error) and zeros (i.e. different states) in the square matrix which correspond to black and white squares in the recurrence plot, respectively. We refer the reader to Marwan et al. (2007) for examples of such recurrence plots. The visual interpretation of recurrence plots requires some experience (Marwan, 2012). For example, the occurrence of white bands in the recurrence plot indicates a disruption or non-stationarity, meaning that the states at these time points are far from the norm. The quantification of this kind of plots offers a more objective measure and caused a surge in popularity among scientists.

At first, these kinds of measures were developed by Zbilut and Webber Jr. (1992) and later extended to incorporate new measures of complexity (Marwan et al., 2002).

1) The first and simplest measure derived from the recurrence plot is the recurrence rate (RR). The recurrence rate reflects the density of the recurrence points, it can be viewed as the mean probability that any state will recur.

2) Determinism (DET). Determinism is the percentage of recurrence points which form diagonal lines. White noise, for example, has a recurrence plot with almost only single dots and very few diagonal lines. This measure quantifies the probability that a recurrence of any state will further recur in the future. This measure corresponds to the segments in phase space that run parallel for some time (Schinkel et al. 2009).
3) Laminarity (LAM). This measure comes down to the percentage of recurrence points which form vertical lines with a certain minimum length. The measure is called laminarity because it reflects the amount of laminar phases in the system. More laminarity in the data may reflect more frequent 'seeds' for synchronised dynamics (Hirata & Aihara, 2011).
4) Averaged diagonal line length (L). The mean length of the diagonal lines. Diagonal lines represent the segments in phase space that run parallel for a certain amount of time. L is a measure for the average time that two segments of the trajectory are close to each other. It is interpreted as the mean prediction time.
5) Longest diagonal line (L_{max}). This measure is related to the exponential divergence of the phase space trajectory (Marwan et al., 2007)
6) Trapping time (TT). This measure indicates the average length of the vertical lines. Vertical lines represent segments which remain in the same phase-space region for some time. The trapping time therefore represents how long the system remains in a specific state such as the length of a transition state. Whereas laminarity reflects the time for the transition to take place (Bosl et al. 2018)
7) Longest vertical line (V_{max}). Analogous to the longest vertical lines in the recurrence plot. However, contrary to the measure based on the diagonal lines, these measures are able to find chaos-chaos transitions (Marwan et al., 2007).
8) Entropy (ENTR). This measure of entropy reflects the complexity of the diagonal lines in the recurrence plot and is related to the sample entropy (see below) (Marwan et al., 2007).

A second nonlinear measure that can be used to characterize any complex dynamic system is the sample entropy (SampE). This measure is derived from approximate entropy, a measure that estimates the amount of regularity or unpredictability of a physiological time signal. Low approximate entropy means there is a high degree of regularity in the sample. SampE is the negative natural logarithm of the probability that two sequences, which are similar for certain amount of time points, remain similar for the next time point. This measure does not include self-matches. The lower the value on the SampE, the more 'self-similarity' in the time series or the less noise is present in the signal. Two main advantages of SampE over approximate entropy are 1) the consistency of SampE compared to approximate entropy and 2) the independence of the length of the data (Richman & Moorman, 2000).

A third and final nonlinear measure that we discuss to characterize the dynamics of the brain signal, is Detrended Fluctuation Analysis (DFA). Formally, this measure is used to characterize the self-affinity of a signal. Basically, this quantifies the long term memory of the signal. It gives an indication if the signal is more, less or equally likely to continue in the same trend as during the previous steps. DFA is related to the Hurst exponent, with the important distinction that DFA can be applied to nonstationary signals (Bosl et al., 2018). However, Bryce and Sprague (2012) show that DFA introduces artifacts for nonlinear trends.

The measures covered above all relate to the dynamic characteristics of the EEG signal and thus to the features of the brain that produced the signal. The nonlinear measures covered are all well established in the field from which they originated (i.e. physics, engineering, electronics, etc), yet their potential and limitations in the framework of neuroscience has only recently begun to be explored. Not only is their potential as a biomarker for autism not fully understood as of yet, the meaning in terms of neural functioning requires more research in the future (Bosl et al., 2018). Some steps in that direction are taken here as we employ these measures in an attempt to distinguish between children with a positive screening for autism (i.e. based on the Q-CHAT administered at 24 months) and a negative screening (i.e. infants with a negative screening using the Q-CHAT). If these EEG-derived measures contain valuable information on brain functioning, it can lead us to predict autism and at the same time study the trajectory of brain development in typically and atypically developing children. Traditional statistical models, like regression analyses, are less suited to capture the brain state based on this high dimensional and noisy data. A method that is better suited for the prediction of which child is susceptible to autism based on these measures, is the use of artificial intelligence. More precisely, the use of machine learning algorithms based on supervised learning.

Prediction & Classification Using Machine Learning

Machine learning algorithms will become increasingly popular as a tool to classify children who are, based on functional measurements of their brain functions, at risk of developing of autism. Machine learning is able to make full use of the multidimensionality and the nonlinearity of the EEG signal and is therefore an ideal candidate for classification based on neurological patterns. Before a classifier can be trained to distinguish between the EEG signal of typically developing infants and those in need of further monitoring, the data must first be transformed.

Pre-processing

The raw EEG signal does not only contain neuronal activity. Physiological noise (e.g. eye blinks, muscle contractions, movement, etc.) and non-physiological noise (e.g. power from the electrical socket, etc.) contaminate the signal and reduce the signal-to-noise ratio (SNR). In general, pre-

processing of an EEG signal entails re-referencing, filtering, epoching, removing bad epochs, interpolating bad channels and removing remaining artefacts using independent component analysis (ICA). A review by Li et al. (2020) offers a general pipeline that can be applied to resting state signals. Although this offers a general guideline, the exact steps and order of the steps should be adjusted based on the nature of the data and the analysis techniques. The first step during pre-processing is, generally speaking, re-referencing. During EEG, we measure the potential for current to flow from one electrode to another, because of this we cannot speak of voltage at one electrode. To measure this potential, we need an active, a reference and a ground electrode. The voltage is measured between the active electrode (A) and the ground electrode (G). The ground electrode, however, always contains noise since it is connected to the internals of the amplifier. For this reason, the voltage between the reference (R) and the ground electrode (G) is measured as well. By subtracting (A – G) – (R – G) = A – R, any noise in common to A and R will be eliminated. One can also choose to re-reference the data offline. For this operation, ideally a reference with a stable and zero potential is chosen. Since this does not exist, a frequently used alternative is the use of an average reference.

Next in the pre-processing pipeline is the filtering. A large part of the artefacts, that are generally present in the EEG signal, have a distinct frequency characteristic. These artefacts can be taken care of by using filters. Four types of filters exist. Low-pass filters attenuate the frequencies that exceed a certain value. This is important given the Nyquist theorem, which states that the highest frequency in the signal should not exceed half of the sampling rate since this can cause aliasing (i.e. high frequency oscillations showing up as low frequencies) and as such generate confounds. High-pass filters, on the other hand, are necessary to reduce confounds originating from the slow drifts of, for example, skin potentials (i.e. the slow opening and closing of skin pores which causes impedance changes). Besides the low-pass and high-pass filter, there are also band-pass filters. Band-pass filters are a combination of the previous two and attenuate the power of frequencies outside the specified range. A last kind of filter is the notch filter. This filter is similar to the band-pass filter with the difference that it suppresses the power in a very narrow frequency band. This last filter is used to suppress the presence of frequencies of electrical devices in the signal. In Europe, the line noise of electrical devices is situated around 50Hz. Filters should be used sparingly since they distort the data and reduce the temporal precision. Extra effort during data collection in order to collect clean data will pay off during the analysis.

A following step that is often conducted, yet not strictly necessary, is the parsing of the data into smaller segments. These segments are often formed by the experimental events, but for resting state data it is often chosen to segment the data in two pieces of two seconds.

The fourth step in the pre-processing pipeline entails removing bad segments and interpolating channels that malfunctioned. While segments containing artefacts can easily be removed without much

data loss, this is not the case for entire channels. Channels can malfunction when they lose contact with the scalp or when two neighbouring electrodes get connected due to excessive gel use. Such bad channels can be replaced by the average of the surrounding electrodes but this has its limits and as stated earlier, there is no substitute for clean data.

A step, that is usually preformed last, is the use of ICA. ICA decomposes the EEG data into components that attempt to identify independent sources of variance in the data. This technique is especially useful to remove eye blinks. Blinks have a large amplitude and originate from a discrete source. Because of this, they are easily picked up using ICA. Some other forms of movement can be removed as well by applying ICA. In general, however, one should be careful when deleting components since many will contain neural activity.

When these steps are completed, we end up with clean (or at least cleaner) data. This is a necessary step before a large corpus of data can be fed to a machine learning algorithm. There is another critical step that needs to be completed before a classifier can be trained to discriminate between patterns of EEG from a typically developing child versus a child requiring follow-up. This step involves 'focussing' the attention of the learning algorithm. We want to reduce the dimensions of the resulting data to those dimensions containing relevant information in order to fully utilize the power of the classifier.

Feature Extraction and Selection

Feature extraction is actually a transformation from data to information. Information that is more suited for classification purposes. Generally, in neuroimaging classification research, feature extraction involves calculating several numerical parameters from the time-series data that can be fed to a suitable classifier. These can range from statistical parameters of the signal to complexity measures to functional connectivity measures. As discussed in the section of endophenotype above, all these potential biomarkers can be used as features.

The dimensions of EEG data are vast. There are often 32 (or more) electrodes, each containing overlapping and specific information. Using these electrodes one can quantify the power in the six distinct frequency bands and within these frequency bands many more features can be identified (i.e. nonlinear measures, see endophenotype section). The dimensionality is too large for the machine learning algorithms to make meaningful predictions. This is known as the curse of dimensionality, with increasing numbers of variables the sample needed to achieve sufficient accuracy rises exponentially. Feature selection is a method of dimensionality reduction that helps in harnessing the full potential of the machine learning algorithm. Reducing the dimensions of a dataset before using machine learning increases the learning accuracy and computational efficiency since it removes redundant and irrelevant data. Using dimensionality reduction techniques also increases the interpretability of the obtained

results and reduces training time. Many feature selection techniques have been developed and all of them have specific strengths and weaknesses. Khalid et al. (2014) offer a review of these strengths and weaknesses.

Feature selection entails the selection of a subset of features, not a transformation. The best subset of data is the one with the smallest number of dimensions that is able to capture the underlying pattern in the data. Feature selection is basically a search problem. Feature selection algorithms can therefore be split up in three distinct elements: the search organisation (i.e. exponential, sequential, random), how the subsets are generated (i.e. forward, backward, compound, weighting and random) and how it evaluates the generated subsets (i.e. divergence, dependence, information, etc). The methods that employ these elements can be categorized as filters, wrappers and hybrid models (Khalid et al., 2014). The different methods trade-off efficiency and reliability. The wrapper methods are more reliable (i.e. they can always provide the best subset of features) and they incorporate the interaction between variables. A downside of this method is the high computing cost since each feature set is evaluated by actually training a model on it (i.e. a machine learning model is 'wrapped' in this method). As a consequence of this they are also prone to overfitting.

Filter methods, on the other hand, are faster and less computationally heavy but they tend to produce less reliable results compared to the wrapper methods. Filter methods quantify the relevance of a feature by measuring the correlation between that feature and the dependent variable. A balance between these methods is struck in so-called hybrid methods. The combination of a filter step and a wrapper step allows this hybrid method to combine the strong points of both methods.

Ladha and Deepa (2011) give an overview of frequently used algorithms for feature selection. Discussing all the possible algorithms is beyond the scope of this thesis. One method that has been used in the classification of EEG-derived features is the recursive feature elimination algorithm (SVM-REF; Bosl et al., 2018). This wrapper method uses backward elimination. The features that produce the largest margin between the datapoints and the separating hyperplane (see classification) are withheld. SVM-REF uses the weight vector as a ranking criterion. The features with the lowest weight (i.e. least significant) are removed on each iteration.

Dimensionality reduction is another process to improve interpretability of data without loss of information. One of the most popular methods is principal component analysis (PCA). PCA implies a linear transformation of the data that simultaneously minimizes redundancy (i.e. covariance) and maximizes the information (i.e. variance) (Cateni et al., 2013). The principle component method does suffer from some limitations. It assumes that the relationship between the variables is linear and that the variables are all scaled at the numeric level. Some methods circumvent these limitations. One example is the nonlinear principle component (Linting et al., 2007). In these nonlinear principle component analysis, variables are quantified during the component extraction and not analysed as they

are (which is the case in PCA). Another method that is applicable, even when the relation between the variables is nonlinear, is the kernel principle component analysis (KPCA). This method uses the kernel trick: the use of a kernel function which enables it to operate in multidimensional feature space and computes the inner product between the images of the data pairs in feature space without the need to calculate the coordinates in this space (see classification).

Once the appropriate number of features from the raw EEG data are extracted and then selected or transformed, depending on the specific demands of the situation, we arrive at the crucial step. The data is now ready for machine learning algorithms to try and classify them as belonging to a typically developing infant or one in need of further monitoring.

Classification

Once the different features are extracted and selected, the next step in the analysis process is the classification. Here, we discuss instances of supervised learning, namely classification problems where the target prediction (i.e. the screening outcome) is known at the time of analysis. In a recent review, Hyde et al. (2019) provide an overview of the different machine learning algorithms used in the research of autism diagnoses. In the articles covered in this review, most studies made use of support vector machines (13 out of 35 studies). Other frequently used methods are alternating decision trees, least absolute shrinkage operator regression, random forest and neural networks. Among the less popular classification algorithms were ridge regression, deep learning, elastic net regression, linear discriminant analysis, logistic regression, decision tree, conditional inference forest, decision stump, flex tree, naïve Bayes and random tree.

Support vector machines (SVM) are used most frequently and have been deployed in a wide variety of classification problems ranging from the classification of individuals based on behavioural assessment to the classification of genes related to ASD. SVM are a form of a supervised learning algorithm. The main purpose of the algorithm is to fit a hyperplane in the multidimensional feature space that separates both groups. The features that were selected in the previous step make up the different axes of this multidimensional space. There might be multiple possible hyperplanes that separate both groups in feature space. The hyperplane that eventually is eventually chosen, is the one that maximizes the distance between the groups it separates. This means that the distance between the nearest datapoints to the hyperplane (i.e. the margin) is maximized and this for both sides. If a hyperplane exists that fulfils this characteristic, than this is called a maximum margin hyperplane and the linear classifier it defines, is a maximum margin classifier. The greater the margin, the better the classifier functions (Hyde et al., 2019). The hyperplane is defined by the kernel. The kernel takes the data as input and transforms it into the required form. Different kinds of kernel functions exist: linear,

nonlinear, polynomial, radial basis function and sigmoid. Without previous knowledge about the data, a Gaussian radial basis function is used (Rohilla, 2018).

Another important concept is generalisation. This refers to how well the model performs on classifying a novel data set after learning. A well-known pitfall to avoid and which is related to the generalisation, is overfitting. Overfitting refers to the situation in which the model not only extracts meaningful data but also every detail and noise present in the training set. This amount of detail and noise is not representative of the pattern that the model needs to extract. The result of overfitting is that the model performs very poorly on new data (Rohilla, 2018). In order to avoid overfitting, a so-called soft margin is used. This means that some of the datapoints are allowed to enter the margin, yet this will still be penalised.

A very useful property of this classification algorithm is that it can be used even when the boundary between the groups is nonlinear. This is done by fitting the hyperplane in a transformed feature space. It uses a kernel function to transform the data to a higher dimension in which the nonlinear data becomes linear and a hyperplane that separates both groups becomes possible. This is known as the kernel trick. Because of this feature, is has been mostly used in applications where the distinction between groups is nonlinear. Another advantage is the fact that SVM are effective even in high dimensional spaces, even in the cases where these dimensions are larger than the sample size. The fact that any kernel function can be specified for the decision function makes the application of SVM extremely adaptable. Note, however, that when the number of features exceeds the number of samples, overfitting is a risk and one should choose a kernel function and regularization term according. Support vector algorithms are available using the Python open source package scikit-learn (Pedregosa et al., 2011).

The predictive power of the resulting features and the model used to classify them is tested using the leave-one-out cross validation. Leave-one-out cross validation is a method where the entire pool of subjects but one is used to train the model and construct the hyperplane that most accurately separates both groups. This resulting hyperplane is then used to classify the left out subject. This process is re-iterated for every subject in the participant pool. Based on the information derived from the group of subjects, a prediction is made for the one subject that was left out. We can then compare the prediction with the known outcome of that subject. Once the model has been trained on each possible split of the data set, all the performance scores of the individual runs are averaged. Performance of a classifier is further specified using sensitivity and specificity. Sensitivity is the ability of the classifier to correctly classify an individual with a positive screening of ASD as needing follow-up (i.e. true positive). The specificity, on the other hand, is the proportion of cases which were correctly identified as coming from an individual with a negative screening for ASD (i.e. true negative). The

positive predictive value (PPV) is an overarching measure indicating the proportion of correct positive predictions in the total amount of positive predictions by the classifier.

Last but not least, the significance of the classification results for each method can then be estimated using, for example, a permutation approach. This approach entails that all the data is shuffled randomly (i.e. feature values, subject labels, etc.). After this permutation, the leave-one-out cross validation is repeated and the accuracy of the model is stored. Next, the data is shuffled again and the validation procedure is reiterated as well. When repeating this procedure multiple times a distribution of the model's accuracy under the null hypothesis can be made. No useful classification can be made when the data is shuffled randomly thus the accuracy of the resulting model depends solely on chance. We can then compare the model's accuracy against this distribution of accuracies under the null hypothesis. If the model's classification is more accurate than 95% of the random classifications then the model's classification is said to be statistically significant at $\alpha = 0.05$.

The overarching goal of this study is to build on the work of Bosl et al. (2018) and evaluate the potential of various features derived from EEG recordings as an early biomarker of autism. These features, derived from an easy and uncostly neuroimaging modality which is ideal to administer on young children, could be of immense value to the field of clinical neuroscience if they provide insights into the early neural dynamics of a developmental disorder like autism. These features, combined with pattern classification algorithms could bridge the gap between the need of early intervention and the currently used clinical tools lacking sensitivity in the younger client population (e.g. ADOS-2, Q-CHAT).

Method

Participants

All the participants in this study were infants between 10 and 14 months of age. Participants were recruited through different paths. Siblings of children with ASD or children with feeding issues were recruited via third party instances such as *Kind en Gezin*, *Tanderuis*, outpatient rehabilitation centres *(i.e. Centra Ambulante Revalidatie (CAR))* and individual psychologists. On the other hand, premature infants were recruited by a researcher at the Centre of Developmental Disorders when parents visited the centre for a planned check-up of their child. All participants were brought to the lab and underwent an EEG recording on two occasions, once at 10 months old and once at 14 months old. At 24 months of age, all children were scored using the ADOS-2 scale and the ten red flags version of the Q-CHAT was filled in based on the parent's report. Based on the results of the Q-CHAT, two distinct groups were formed, a group with a positive screening and a group with a negative screening at 24 months of age. Since all these participants had a higher genetic risk of autism, a higher proportion of the sample would likely develop ASD, making it a better sample to test the potential of machine learning algorithms to classify individuals at risk of developing ASD based on their EEG derived features at an early age. In total, 64 infants with a predisposition for ASD participated in this study. The sample consisted of 31 male infants (i.e. 15 preterm (P), 14 with diagnosed sibling (S), 2 with feeding problems (F)) and 33 female infants (i.e. 15 P, 16 S, 2 F). Not all subjects participated in all the scheduled assessments due to Covid-19 measures preventing, amongst others, the scoring on the Q-CHAT at 24 months. The study was approved by the Medical Ethics Commission at Ghent University Hospital following consultations with each participating centre, registration number: B670201733369. All parents read and signed an informed consent form.

Material

All the resting state EEG recordings were made using BrainVision recorder software (Brain Products GmbH, Gilching, Germany). This study used a BrainVision actiCHamp Plus amplifier. Before the onset of recording, the participants' skull circumference was measured and an appropriate cap was fitted using 31 electrodes and the 10/20 layout. During the recording of scalp voltages the Fz electrode was treated as the reference electrode. No external electrodes were used to monitor eye-movements or other physiological responses. Visual inspection of the raw signal was used to assess whether the electrodes were functioning properly.

Procedure

Participants were brought to the lab and written informed consent was obtained from the parents. After fitting the appropriate cap and making sure all electrode impedances were acceptable,

the participants were than taken to a dimly lit, electrical- and sound shielded testing room. While resting state recording in adults usually entails asking them to close their eyes and sit still during the recording, this is not feasible when recording the brain activity of infants. Therefore, the participants were seated on the lap of a parent and were presented short videoclips. The videoclips used in this study were developed by the British Autism Study of Infant Siblings (BASIS) Team, a collaborative research network for infants at elevated likelihood of autism based in the United Kingdom. Two types of videos were used in this study, namely videos showing social interactions (e.g. a woman singing nursery songs) and videos showing non-social toy videos. Every block consisted of 131 seconds. Of this duration, 70 seconds were dedicated to videos showing social interactions and 61 seconds of non-social videos. In total six blocks of these clips were presented to the participants (three social and three non-social). The social and non-social clips were always shown in succession with a visual event related potential (ERP) task separating two blocks.

At 24 months the participants and their parents were invited to complete two additional assessments. Namely, the ADOS-2 diagnostic observation schedule as well as the Q-CHAT, a screening instrument filled in by the parents.

Analysis

First, scores were calculated for the two assessments performed at infants of 24 months old. The scores on the items of the ADOS-2 were converted to algorithm scores based on the instructions in the manual. In the analyses, the ADOS-2 scores were used as a continuous measure. For the Q-CHAT, the ten most discriminating items (i.e. 'red flags', items 1,2,5,6,9,10,15,17,19,25; Allison et al., 2012) were used to form a preliminary assessment (i.e. positive or negative screening). The score on these red flag items was summed and a score equal to or greater than three was taken as a positive screening.

For the pre-processing of the raw EEG data, BrainVision Analyzer (Version 2.2.0, Brain Products GmbH, Gilching, Germany) was used. The data was segmented in three ways, social only, non-social only and social and non-social combined. The latter was used for further processing. Before proceeding, the data was high pass filtered using a 1Hz filter and low pass filtered using a 50Hz notch filter. A down-sampling procedure was used to bring the sampling rate down to 512Hz. Based on visual inspection, bad channels were excluded from further analysis. The online reference was the Fz electrode which was later replaced by an average of all electrodes. Once the re-referencing was complete, the data was segmented in 1s epochs. A semi-automated artefact rejection was applied using the following criteria: an epoch was rejected if it exceeded an amplitude of 150 µV or -150 µV, if it showed low activity (which was defined as 0.5 µV max - min) for at least 100ms, if the absolute difference of values in an epoch was greater than 200 µV or if it showed a steep gradient of 50 µV/ms or more. Channels containing more than 50% rejected epochs were interpolated and removed from the re-reference. After pre-

processing the signal was exported as an continuous signal. Using a daubechies-4 wavelet transform the signal was decomposed in the most commonly used frequency bands (and analogous to Bosl et al., 2018): delta (1 - 4Hz), theta (4 - 8Hz), alpha (8 - 16Hz), beta (16 - 32Hz), gamma (32 - 64Hz), high gamma (64 - 128Hz).

Next, the features needed for the classification were extracted from the continuous and pre-processed EEG-signal. Using the frequency bands (defined in the previous paragraph) and for each electrode that was kept after the pre-processing, the logarithm of the root mean square (i.e. the frequency power) in each frequency band and complex system parameters or invariant measures were computed using MATLAB (R2021b). This resulted in five features per frequency band and for each electrode, namely the logarithm of the root mean square (LRMS), the recurrence rate (RR), determinism (DET), entropy (ENTR) and averaged diagonal line length (L).

Once the features were extracted, machine learning algorithms could be used for prediction and classification purposes. Since only five features were extracted from the EEG signal, all the features could be used in the subsequent classification. The following classification analyses were conducted separately on the features extracted from EEG-recordings at 10 months and those at 14 months. First, for the actual computation of the classification, SVM with a radial basis function were used. The SVM was fitted using the default parameters, except for the regularization parameter 'C' which was set to 10. All features (i.e. 5) of all frequency bands (i.e. 6) and using the complete set of electrodes (i.e. 31) were used to train the model. The screening results of the Q-CHAT at 24 months (i.e. positive screening or negative screening) were used as labels. The validation of the prediction of either high likelihood or low likelihood of ASD by the model was based on a leave-one-out cross validation procedure combined with a non-parametric technique for estimation of statistical significance, namely a permutation approach. In a leave-one-out cross validation procedure, the classifier, here a SVM, is trained on all subjects but one to find the separating hyperplane. The resulting hyperplane is then used to classify the one subject that was left out. By repeating this process for all participants, the accuracy of the classifier can be determined. Based on this validation scheme, sensitivity, specificity and positive predictive value were computed of the model's classifications.

To test whether the model's classification performance is statistically significant, a non-parametric technique for estimating significance will be used, namely a permutation approach (Golland & Fischl, 2003). Here, 1000 permutations were used to construct the null distribution and the significance threshold was set at 5% (i.e. $\alpha = 0.05$).

Using one's EEG features, we can go beyond mere classification. We can also attempt to estimate the severity of the ASD symptoms using support vector regression (SVR). In this second step, again all features were used, combined with the score on the ADOS-2 observation scheme as outcome. Using the continuous ADOS-2 scores as measured at 24 months, we could evaluate the predictive

validity of the model by again using a leave-one-out cross validation scheme. Next, we correlated the model's prediction of the ADOS-2 score, based on the nonlinear features derived from EEG recordings at 10 and 14 months of age, to the actual scores of the ADOS-2 diagnostic tool.

As a last classification analysis, the difference scores of the features at 14 months and those at 10 months (i.e. 14M – 10M) were calculated and the previous analyses were repeated (i.e. support vector classification and support vector regression). This is based on the rationale that the developmental trend of the nonlinear features could be predictive of emerging ASD symptoms. The machine learning calculations were all done using the scikit-learn package in Python (Pedregosa et al., 2011).

Finally, a large scale comparison was made between the EEG-derived features of children with a positive screening on the Q-CHAT and those with a negative screening on the Q-CHAT evaluation at 24 months. For both the recordings at 10 months of age and those at 14 months of age an independent samples t-test was used to compare the features of the group with a positive screening and those with a negative screening. This comparison was made for all electrodes (i.e. 31) and all frequency bands (i.e. delta – high gamma). Since this entails a large number of comparisons, the significance threshold is adjusted using the Holm-Bonferroni method (Abdi, 2010). The code used for the feature extraction and the classification is available on: https://github.ugent.be/sboeve/MasterThesis.git.

Results

Due to the nature of the population under study, young children at elevated risk of autism, motion artifacts were ubiquitous in the EEG-recordings. For a small subset of participants, almost all the channels had more than 50% of bad epochs. The data of those participants was excluded from further analysis. This resulted in a total of 54 useful recordings when the participating infants were 10 months old (26 males (12 P, 13 S, 1 F), 28 females (15 P, 13 S); out of a total of 60 EEG recordings) and 40 useful recordings when participants were 14 months old (19 males (10 P, 7 S, 2 F), 21 females (9 P, 10 S, 2 F); out of a total of 47 EEG recordings).

For these participants, five features of their EEG recordings were extracted: logarithm of the root mean square (LRMS), the recurrence rate (RR), determinism (DET), entropy (ENTR) and averaged diagonal line length (L). These features were calculated for each frequency band and for each electrode resulting in a total of 930 features (i.e. 5 features x 6 frequency bands x 31 electrodes) for each participant.

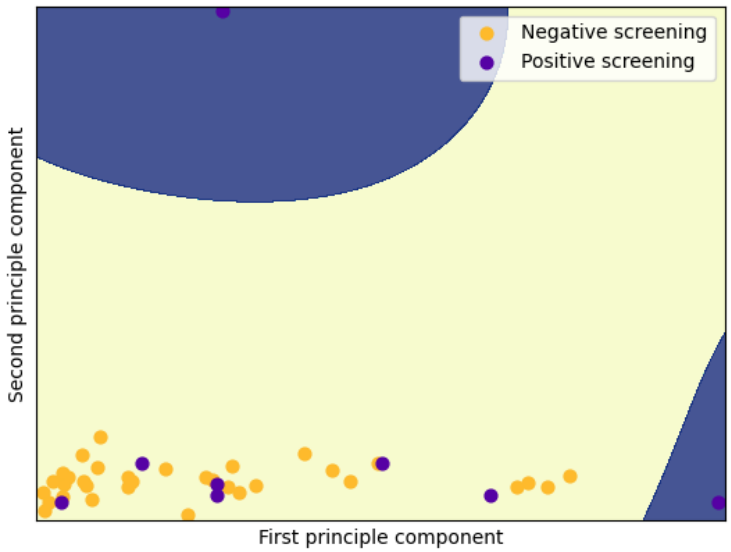
Not all participants underwent the screening at 24 months, this resulted in some missing data points. For the participants with a recording at 10 months old (i.e. 54), 41 participants also had scores on the Q-CHAT of which 33 had a negative screening and 8 had a positive screening. For 26 participants both the EEG recording at 10 months as the ADOS-2 score were available. Out of the total of 40 participants who had an EEG recording at 14 months, 37 had a score on the Q-CHAT (i.e. 33 negative screenings, 4 positive screenings) and 20 had a score on the ADOS-2. There was a total of 28 participants with an EEG recording at both ages and a Q-CHAT score at 24 months (i.e. 25 negative screenings, 3 positive screenings), while only 11 had both EEG recordings and the ADOS-2 score.

Classification Using EEG Features at 10 Months

The outcomes of 41 subjects on the Q-CHAT were predicted based on their EEG features at 10 months of age. Using a SVM and a leave-one-out cross validation procedure, the outcome of a subject was predicted using her or his EEG features at 10 months and a SVM trained on the EEG features of all other subjects. Figure 1 illustrates the decision boundary of the SVM trained on the EEG features of the 10 month old participants.

Figure 1

Decision Boundary of the SVM Trained on EEG Features at 10 Months Old



Note: Using principal component analysis (PCA) the 930 features (i.e. 5 features x 6 frequency bands x 31 electrodes) are reduced to two dimensions for visualisation purposes. Yellow dots represent participants with a negative screening at 24 months using the Q-CHAT. Purple dots represent participants with a positive screening on the Q-CHAT. The blue surface represents the decision boundary of what is classified by the SVM as participants in need of further monitoring (i.e. positive screening). The actual classifier was trained on all 930 features and evaluated using a leave-one-out cross validation procedure.

Using the leave-one-out cross validation procedure, it was determined that the classifier had an accuracy of 78.05% in classifying the Q-CHAT outcome based on the EEG features. The model reached a sensitivity of .0 and a specificity of .97, resulting in a PPV of .0 in the classification of participants as needing further monitoring (e.g. positive screening) or not. This is due to the fact that the model almost always predicted a negative screening based on the EEG features. Table 1 gives an overview of the classifier's performance using EEG features of different ages. The outcome of the non-parametric test using 1000 permutations showed that the model's performance on a random configuration of feature values ranged between 75.60% and 82.93%, with a mean accuracy of 80.43%. Based on this distribution the classifier is situated in the 76th percentile, hence the classification accuracy is not statistically significant. Due to the small dataset and the limited number of positive cases, the model is unable to extract a meaningful pattern from the data.

Table 1

Predictive Classification of the 24 Month Outcome Based on EEG Feature Classification

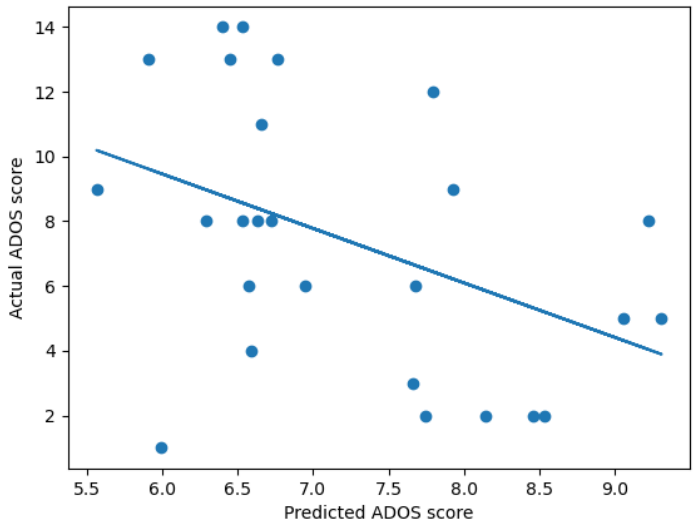
EEG recording	*N* subjects		QCHAT	SVM prediction		Accuracy		
	pos.	neg.		pos.	neg.	Sens.	Spec.	PPV
10 months	8	33	pos.	0	8	.0	.97	.0
			neg.	1	32			
14 months	4	33	pos.	0	4	.0	1	.0
			neg.	0	33			
(14) - (10)	3	25	pos.	0	3	.0	1	.0
			neg.	0	25			

Note: pos. = positive screening, neg. = negative screening, Sens. = sensitivity, Spec. = specificity, PPV = positive predictive value. Age of testing refers to the age at which the EEG recording was made. *N* subjects shows the amount of participants with both a screening at 24 months using the Q-CHAT and a recording at the specified age split over those with a positive screening at 24 months and those with a negative screening. (14) – (10) refers to the predictions based on the difference score of the features for participants who had an EEG recording at both ages. Sensitivity, specificity and positive predictive value were computed according to the standard definition.

Next, we used SVR in order to predict the ADOS-2 scores at 24 months using the EEG features at 10 months. The accuracy of the model was assessed using a leave-one-out cross validation. The resulting predictions of the ADOS-2 scores were then correlated with the actual ADOS-2 scores at 24 months using Pearson's correlation coefficient (see fig. 2). The predicted scores showed a negative correlation with the actual ADOS-2 scores, $r = -.43$, $p = .027$, $N = 26$. The correlation as well as the accompanying *p*-value should be placed in perspective as these are based on only 26 participants.

Figure 2

Correlation of the Predicted ADOS-2 Scores Using EEG Features at 10 Months and the ADOS-2 Score as Assessed at 24 Months

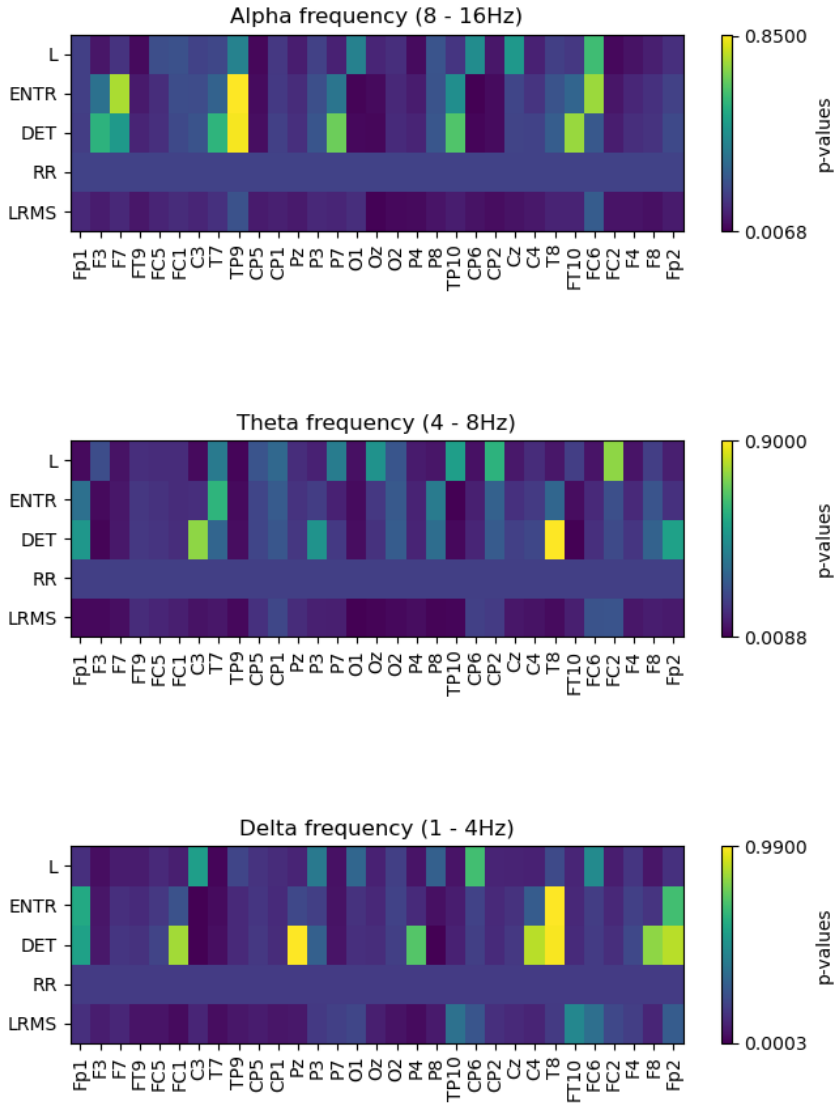


Note: Correlation of the predicted ADOS-2 scores using SVR and the ADOS-2 scores as assessed by a clinician when the infants were 24 months old. The prediction of the ADOS-2 scores was made using EEG-derived features from recordings of infants at 10 months old.

Finally, multiple independent sample t-tests were conducted, comparing the EEG features at 10 months old for the group with a positive screening at 24 months based on the Q-CHAT (N = 8) and the group with a negative screening (N = 33; see fig. 3 – 4). Since this entails multiple comparisons, the p-values were corrected using the Holm-Bonferroni method with α = 0.05. The values of all 930 features at 10 months were compared between the aforementioned groups (i.e. positive vs. negative screening at 24 months) but none of the tests reached statistical significance. There were three instances in which the group comparison on the EEG features showed a trend towards significance. All three were situated in the delta band, namely for DET and ENTR at electrode C3 and DET at electrode P8. The p-values and the corrected significance thresholds (i.e. α) are as follows: Delta, DET, C3: Δ = 0.0081, $t(39) = 3.96$, p = 0.00031, α = 5.38e-05; Delta, ENTR, C3: Δ = 0.031, $t(39) = 3.92$, p = 0.00034, α = 5.38e-05; Delta, DET, P8: Δ = 0.0061, $t(39) = 3.40$, p = 0.0016, α = 5.39e-05. Δ refers to the difference of the means of the feature in the group with a positive and a negative screening. A positive difference means that the feature had, on average, a higher value in the group with a negative screening. It is noted that the RR values for all features and frequency bands is the same for all participants within a group of the same outcome. The cause of this artifact is unknown and will be investigated further.

Figure 3

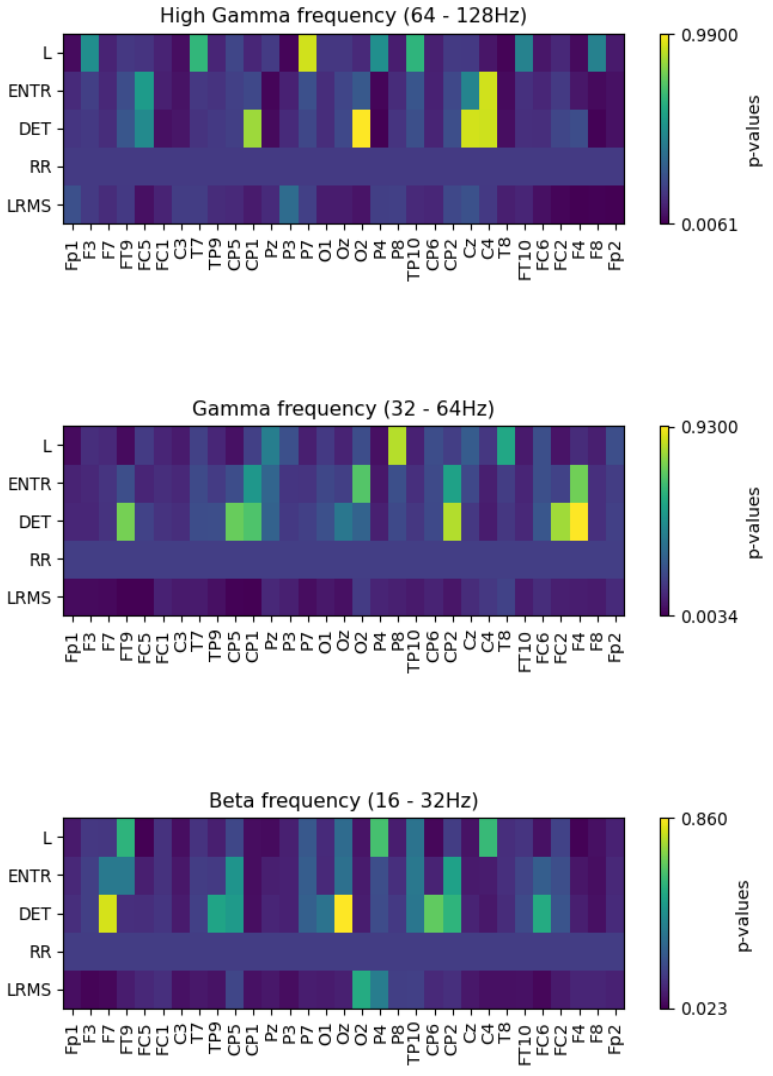
Comparison of the Values on the EEG Features Between the Group With a Positive and the Group With a Negative Screening – EEG Recording at 10 Months



Note: Comparison of the values of the features for the delta, theta and alpha frequency band between the group with a positive screening (*N* = 8) and the group with a negative screening (*N* = 33) at 24 months. The vertical axis of each subplot represents the features extracted from the EEG signal recorded at 10 months. The horizontal axis shows the electrode site. The colours indicate the *p*-value of the comparison of the values of those with a positive screening and those with a negative screening. Uncorrected *p*-values are shown. Note that the range of *p*-values is slightly different for each subplot.

Figure 4

Comparison of the Values on the EEG Features Between the Group With a Positive and the Group With a Negative Screening – EEG Recording at 10 Months



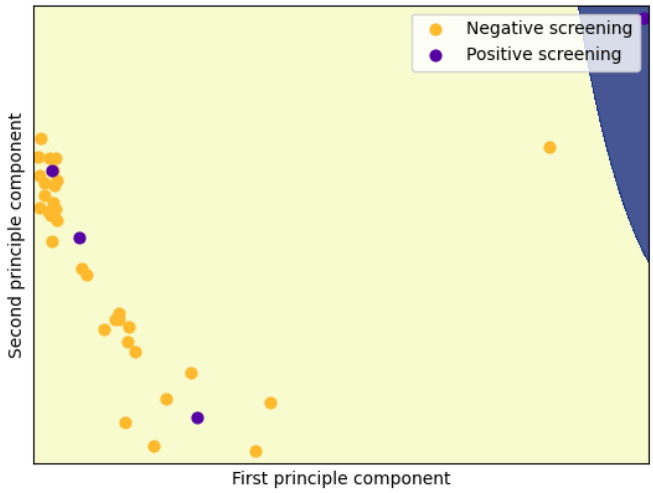
Note: Comparison of the values of the features for the beta, gamma and high gamma frequency band between the group with a positive screening (*N* = 8) and the group with a negative screening (*N* = 33) at 24 months. The vertical axis of each subplot represents the features extracted from the EEG signal recorded at 10 months. The horizontal axis shows the electrode site. The colours indicate the *p*-value of the comparison of the values of those with a positive screening and those with a negative screening. Uncorrected *p*-values are shown. Note that the range of *p*-values is slightly different for each subplot.

Classification Using EEG Features at 14 Months

The same procedure was used for the classification of the EEG features at 14 months of age. For 37 participants both the EEG features at 14 months and the Q-CHAT score at 24 months were available (i.e. 33 negative, 4 positive screenings). Figure 5 shows the learned decision boundary by the SVM using the EEG features at 14 months. Using a leave-one-out cross validation, the SVM was able to classify the EEG features at 14 months with an accuracy of 89.19%. The classifier had a sensitivity of .0 and a specificity of 1, resulting in a PPV of .0 in the classification of EEG features at 14 months as belonging to a participant with a positive or a negative screening for ASD at 24 months (see table 1). This low sensitivity is due to the classifier classifying all instances of the EEG features as belonging to a participant with a negative screening at 24 months. The outcome of the non-parametric test using 1000 permutations showed that the model's performance on a random configuration of feature values ranged between 81.08% and 91.89%, with a mean accuracy of 88.44%. The classifier's accuracies places it in the 84th percentile, hence not statistically significant.

Figure 5

Decision Boundary of the SVM Trained on EEG Features at 14 Months Old



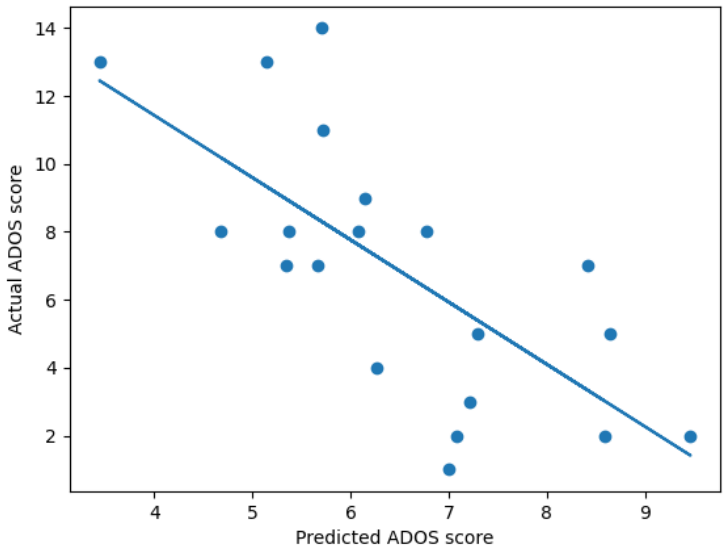
Note: Using principal component analysis (PCA) the 930 features (i.e. 5 features x 6 frequency bands x 31 electrodes) are reduced to two dimensions for visualisation purposes. Yellow dots represent participants with a negative screening at 24 months using the Q-CHAT. Purple dots represent participants with a positive screening on the Q-CHAT. The blue surface represents the decision boundary of what is classified by the SVM as participants needing further monitoring (i.e. positive screening). The actual classifier was trained on all 930 features and evaluated using a leave-one-out cross validation procedure.

Analogous to the analysis of the EEG features at 10 months, we used SVR in order to predict the ADOS-2 scores at 24 months using the EEG features at 14 months. Using a leave-one-out cross

validation method, the classifier's predictions were assessed and in a next step correlated with the actual ADOS-2 scores as measured at 24 months. The outcome was a strong negative correlation between the prediction and the outcome, as can be seen in figure 6. Indicated by a strong negative Pearson's correlation coefficient, r = -0.70, p = 0.00052, N = 20, the model was not able to predict the degree of autistic symptoms at 24 months as indexed by the ADOS-2 scores. It has to be noted again that this correlation is based on a very limited number of participants and should therefore be interpreted with caution.

Figure 6

Correlation of the Predicted ADOS-2 Scores Using EEG Features at 14 Months and the ADOS-2 Score as Assessed at 24 Months



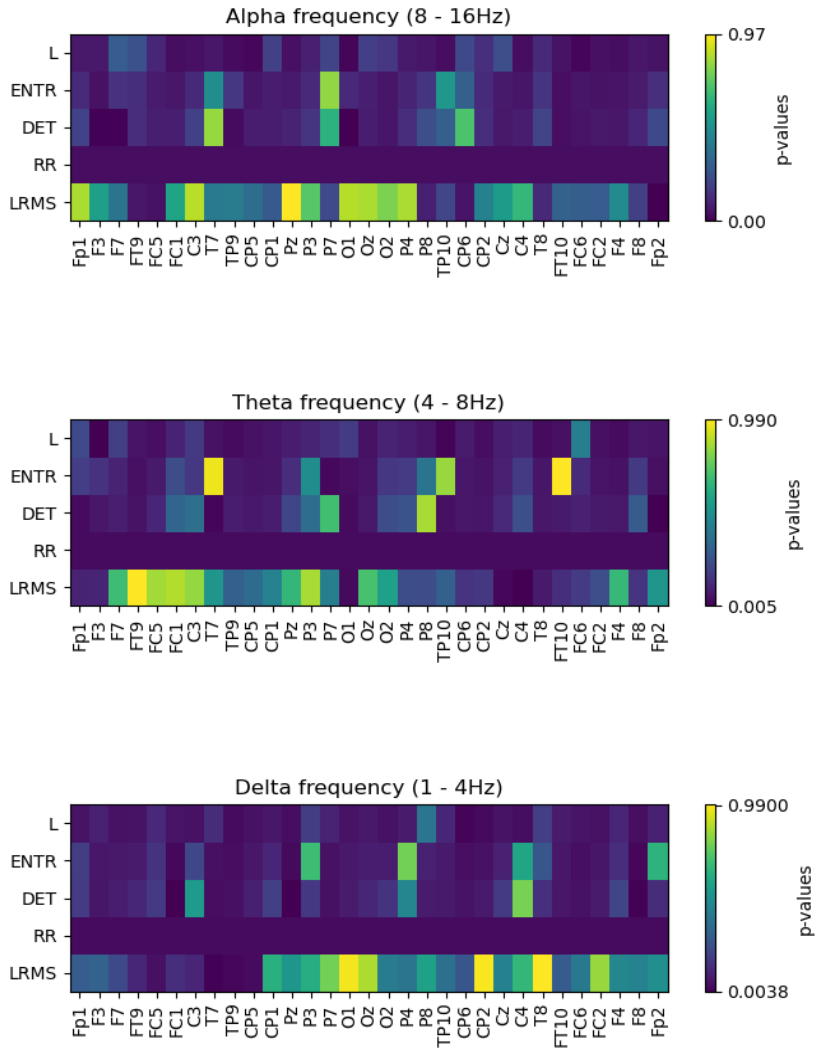
Note: Correlation of the predicted ADOS-2 scores using support vector regression and the ADOS-2 scores as assessed by a clinician when the infants were 24 months old. The prediction of the ADOS-2 scores was made using EEG-derived features from recordings of infants at 14 months old.

As a last step in the analysis of the EEG features at 14 months, the EEG features of the group with a positive screening (N = 4) and the group with a negative screening (N = 33) were compared (see fig. 7 – 8). For this, multiple independent sample t-tests were used, combined with a Holm-Bonferroni correction with α = 0.05. None of the tests comparing the values of the EEG features between both groups reached statistical significance. There were again some instances in which a trend towards significance could be observed. This was the case for DET at electrode O1 in the alpha frequency band, DET at electrode Oz in the gamma frequency and LRMS at Oz in the gamma frequency band as well. The p-values and the corrected significance thresholds are as follows: alpha, DET, O1: Δ = -0.014, t(35) = -4.45, p = 8.30e-05, α = 5.38e-05; gamma, DET, Oz: Δ = -0.010, t(35) = -3.82, p = 0.00052, α = 5.38e-05; gamma, LRMS, Oz: Δ = 0.32, t(35) = 3.37, p = 0.0018, α = 5.39e-05. Δ refers to the difference of the

values of the feature in the group with a positive and the group with a negative screening. A positive difference means that the feature had on average a higher value in the group with a negative screening. Here the same artifact for RR is present as before and will be investigated further.

Figure 7

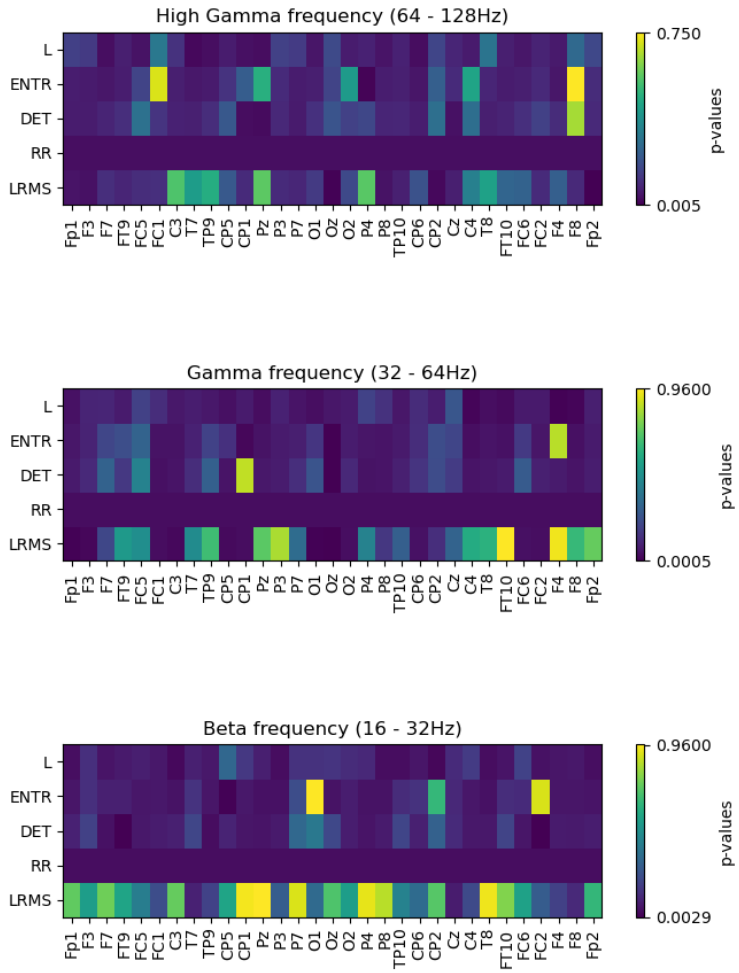
Comparison of the values on the EEG features between the group with a positive screening and the group with a negative screening – EEG recording at 14 months



Note: Comparison of the values on the features for the delta, theta and alpha frequency band between the group with a positive screening (N = 4) and the group with a negative screening (N = 33) at 24 months. The vertical axis of each subplot represents the features extracted from the EEG signal recorded at 14 months. The horizontal axis shows the electrode site. The colours indicate the p-value of the comparison of the values of those with a positive screening and those with a negative screening. Uncorrected p-values are shown. Note that the range of p-values is slightly different for each subplot.

Figure 8

Comparison of the Values on the EEG Features Between the Group With a Positive Screening and the Group With a Negative Screening – EEG Recording at 14 Months



Note: Comparison of the values on the features for the beta, gamma and high gamma frequency band between the group with a positive screening (N = 4) and the group with a negative screening (N = 33) at 24 months. The vertical axis of each subplot represents the features extracted from the EEG signal recorded at 14 months. The horizontal axis shows the electrode site. The colours indicate the p-value of the comparison of the values of those with a positive screening and those with a negative screening. Uncorrected p-values are shown. Note that the range of p-values is slightly different for each subplot.

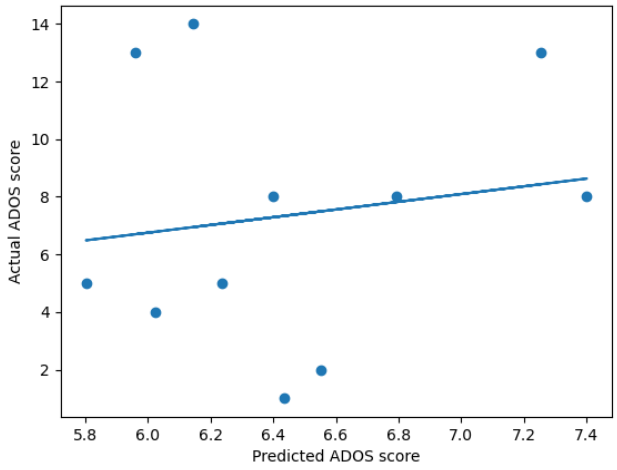
Classification Using the Difference Score of the EEG Features

Grounded in the rationale that the developmental trend of the nonlinear EEG features could be indicative of emerging ASD symptoms, a final classification was made using the difference score of the nonlinear features (14 months – 10 months). First, the participants were selected of which an EEG recording was available at both 10 and 14 months and who in addition had a screening using the Q-CHAT at 24 months. There were 28 participants who fulfilled these criteria. Next, the ability of a SVM to classify participants based on a difference score of the EEG features was assessed. Since there was little data available, the leave-one-out cross validation method was the most appropriate to probe the model's prediction accuracy. The model trained on the difference scores classified the left-out participants with an overall accuracy of 89.29%. The model had a sensitivity of .0 and a specificity of 1. The positive predictive value was .0 (see table 1). The outcome of the non-parametric test using 1000 permutations showed that the model's performance on a random configuration of feature values ranged between 78.57% and 92.86%, with a mean accuracy of 88.78%. Based on this distribution the classifier is situated in the 82th percentile, hence the classification accuracy is not statistically significant.

Using an analogous approach as before, a prediction of ADOS-2 scores at 24 months based on the differences scores was made by means of SVR. Figure 9 shows the correlation of the predicted ADOS-2 scores and the actual outcome. The Pearson's correlation coefficient was .15, yet this was not significant at $\alpha = 0.05$, $p = 0.65$. This correlation is based on a very limited number of participants (N = 11) and should be interpreted with caution.

Figure 9

Correlation of the ADOS-2 Score as Assessed and the Predicted ADOS-2 Scores Using the Difference Score of the EEG Features Derived at 14 Months and Those Derived at 10 Months



Note: Correlation of the predicted ADOS-2 scores using support vector regression and the ADOS-2 scores as assessed by a clinician when the infants were 24 months old. The prediction of the ADOS-2 scores was made using the difference score of the EEG features at 14 months and at 10 months (14 months – 10 months).

Discussion

The primary goal of this study was to explore and evaluate the potential of EEG features, both simple and nonlinear, as an early biomarker of emerging autism. Since EEG is relatively easy to administer and comes at a modest cost, it could have a great impact in the early detection and intervention of autism if the information captured by this modality can be successfully combined with pattern classification algorithms.

The results presented above paint a rather grim picture with respect to this goal as the pattern classification algorithms used here (i.e. SVM and SVR) were not able to discern meaningful patterns in the nonlinear EEG features that can be used to classify infants at risk of autism. All classifiers showed high accuracy using the leave-one-out validation method but only due to the fact that it classified almost every instance as a negative case (i.e. typical likelihood of autism). This led to the classifiers having a high specificity (i.e. the true negative rate; here approximately 1) but a very low sensitivity (i.e. the true positive rate; here .0) and positive predictive value (i.e. the proportion of true positives in the total amount of positive predictions; here .0). The performance of the support vector regression in predicting the ADOS-2 score at 24 months was very poor, with mostly strong negative correlations between the predicted and the actual score.

These results are not surprising given the small sample size that was available to train the classifier. The dataset available to train the classifier on was smaller than anticipated. This was due to the Covid-19 measures preventing the administration of the ADOS-2 and Q-CHAT at 24 months. This issue will be alleviated for the most part once the remaining participants can be scored on the Q-CHAT and the ADOS-2. On top of that, there were very little participants with a positive screening on the Q-CHAT at 24 months, leaving insufficient variance for the classifier to learn to differentiate the EEG features of those with a positive screening and those with a negative screening. The scarce amount of data is the main shortcoming of the present study. One potential measure to accommodate for this flaw directly, is to loosen the stringent pre-processing step of excluding participants for which more than 50% of the epochs had to be rejected. Here, it caused 13 participants to be excluded from further analyses (i.e. 6 recordings at 10 months old, 7 recordings at 14 months old). Lowering this strict criterion in this case would have had little effect since most participants which were rejected had overall very poor quality EEG data (i.e. only 10% or less of the epochs were kept after pre-processing). In their paper, Bosl et al. (2018) did not perform any selection based on visual inspection. While this reduces the pre-processing time and the number of exclusions, it also has potential side effects. Subtle behavioural characteristics associated with ASD might introduce differences into the EEG signal not related to brain dynamics when such motion artifacts are not rejected.

Next, the labels of the participants used in this study should be put in perspective. All participants included in this study had an elevated likelihood of autism (i.e. siblings with autism, premature infants or feeding issues). The results of the screenings at 24 months were used as labels to train the classifier. It must be noted that these screenings are not clinical diagnoses. The screener used here, Q-CHAT, has its limitations. The outcome of the screener does not necessarily reflect the eventual clinical diagnosis. The labels used here were the second best option as the best estimate diagnoses were not available yet. This also touches upon a related issue, namely all participants included here had an elevated likelihood of developing autism. It is possible that this caused the brain dynamics of those with a positive and those with a negative screening of autism to be more similar than would have been the case between those with a low likelihood of autism (i.e. no diagnosed sibling and full term) and a negative screening and those with a high likelihood (i.e. diagnosed sibling or preterm) and a positive screening. Future studies should strive to include a low likelihood control group if this is feasible.

The comparison of the EEG features of both groups (i.e. those with a positive screening and those with a negative screening) yielded little insight into the brain dynamics of those with suspected autism. No significant differences were found in the values of the EEG features when comparing the group who received a positive screening at 24 months and the group who received a negative screening. This was the case for both the EEG features at 10 months old and at 14 months old. These observations appear to be in line with the results reported in Bosl et al. (2018). There, an apparent shift in EEG features was visible at an age of approximately 12 months. For RR, L and DET the values for the ASD group were higher than those of the low risk controls before the 12 month mark and lower for the ASD group beyond 12 months old. The ENTR feature showed an opposite trajectory as it was lower for infants at high risk of autism before 12 months of age and higher afterwards. The EEG recordings in this study all took place in close temporal proximity to this developmental 'shift' observed by Bosl et al. (2018). This could potentially explain the lack of significant differences. However, some of the comparisons displayed a trend towards significance. In our study, for the EEG features at 10 months old, the group with the positive screening showed lower values on DET and ENTR for central and parietal electrodes. While these features and scalp location combinations showed no significant differences in Bosl et al. (2018), they displayed a opposite pattern as well, since the ASD group scored higher on those features.

In our study, the results at 14 months showed a trend towards significance for DET at the occipital electrodes in the alpha (8 – 16 Hz) and gamma (32 – 64 Hz) band. For these features, the group with a positive screening scored higher. The same pattern can be observed in the results of Bosl et al. (2018), in which the corresponding comparison also did not reach significance. Additionally, there is a similarity between the current results and the results in Tierney et al. (2012). In the current study, the comparison at 14 months showed that the group with the positive screening had lower gamma

power at the occipital electrodes compared to the group with the negative screening. This fits the results reported by Tierney et al. (2012) as they observed lower gamma power in infants at high risk of autism from 6 to 24 months. Roughly speaking, there are some communalities while on the whole the observed patterns do not match, making it hard to interpret the results of the current study.

This type of inconsistency between studies is not rare in studies investigating neural correlates or endophenotypes of ASD. Gurau et al. (2017) reviewed the utility of three different methods of EEG signal analysis in diagnosing ASD and delineating the subtypes. These authors identified 40 papers on ASD diagnosis using either functional connectivity analysis, power analysis or a comparison of information dynamics. The results are striking, as all studies covered in this review established significant differences between ASD and non-ASD subjects on these measures. On the other hand, the heterogeneity in these results prevented the derivation of general patterns. This led the authors to conclude that none of the methods alone are currently suited as a diagnostic tool or indicator that a child needs further evaluation. Their analysis of the current state of the field led them to provide three main areas of improvement for future studies. The field needs better experimental designs (preferably longitudinal designs), more advanced analysis methods and more data in general. While these are solid and reasonable recommendations, it neglects the main source of heterogeneity in the results. Besides wonky designs and a lack of subjects, the diagnosis of autism itself prevents the generalisability of these types of studies.

In psychology and psychiatry there has been a long tradition of conceptualizing mental disorders using a number of behavioural symptoms and the degree of impairment the subject experiences. This way of defining mental illness has numerous benefits as it provides clear cut diagnostic categories, it is reliable and makes the diagnosis relatively easy. However, this approach has many downsides.

First and foremost is the heterogeneity of people that qualify for a certain diagnosis. To qualify for ASD one must display persistent deficits in each of the three areas of social communication (i.e. deficits in social and emotional reciprocity, in non-verbal communication and in the establishment and maintenance of social relations) and at least two of four types of restricted repetitive behaviours (i.e. repetitive movements, excessive adherence to routines, highly specific interests and hypo – or hyperreactivity to sensory experiences; APA, 2013). Two people diagnosed with autism can thus manifest a widely different array of atypical behaviours despite having the same diagnosis. Although this is not problematic in itself, it hinders researchers in pinpointing the specific neurobiological aspects of the disorder as these might differ among patients receiving the same diagnosis.

Second, people suffering from one sort of mental disorder tend to meet the criteria for other disorders as well (i.e. comorbidity). This has led to the concern that the current taxonomy of mental disorders does not reflect the underlying neurobiological aspects of the disorders. A final major concern

is that the criteria for defining a mental disorder are created by consensus among clinical practitioners. As a result, the criteria are always somewhat arbitrary and the similarity between those who just meet and just miss the criteria is very large (Doernberg & Hollander, 2016). Due to these concerns there is a growing realisation that it is important to consider dimensional conceptualisations.

This realisation has led to the launch of the Research Domain Criteria (RDoC) project by the National Institute of Mental Health (NIMH, 2009). The RDoC project's principle is that, in order to understand mental disorders and provide adequate care for patients, a comprehensive account of typical and atypical brain maturation is indispensable. For this project, scientists and practitioners adhere to a research strategy which currently focusses on six major domains of human cognitive functioning. These domains are negative valence, positive valence, cognitive systems, systems for social processes, arousal/regulatory systems and sensorimotor systems. Each of these domains consists of constructs (e.g. acute threat, potential threat, etc. for negative valence systems) that can then be studied using different units of analysis such as self-report measures, behavioural measures, physiological measures, genes, etc. The goal is to provide insights into the basic cognitive and behavioural processes and how they impact mental health. Such knowledge is critical in order to improve the diagnosis of mental health issues, inform future versions of mental disorders nosology and research on potential treatments.

How could RDoC inspired studies inform the research on EEG classification of people suffering from autism? First, by conducting studies focussing on psychological constructs relevant to the psychopathology at large rather than merely examining a mental disorder category. We need to understand the individual symptoms patients are confronted with rather than focussing on the collection of symptoms ascribed to a certain diagnosis. For research on ASD this would involve, amongst others, studying the RDoC construct of motor actions with subconstructs such as action planning and selection, sensorimotor dynamics and inhibition or termination of motor plans since repetitive motor movements and echolalia are frequently reported symptoms in the context of ASD (e.g. Harrison et al., 2021; Tschida & Yerys, 2021). This approach has the potential to yield a more clear understanding of the deficits that patients are confronted with. Focussing on specific diagnoses can obscure results due to the inherent heterogeneity of diagnoses based on DSM-5 criteria.

Second, by integrating the assessment of behavioural and biological measurements in a more specific context as compared to resting state. Take as an example the RDoC domain of social processes. This domain refers to all functions mediating responses in an interpersonal setting. A domain for which people suffering from ASD are thought to have persistent deficits. One of the constructs of the domain of social processes is social communication with subconstructs such as reception and production of facial communication etc. Studies investigating the neural correlates of the perception of emotion through facial communication in typical and atypical functioning brains could provide valuable insights

(e.g. Clarkson et al., 2019; Hennessey et al., 2018; McVey, 2019). For example, such studies could broaden our understanding of the neurobiological function and meaning of some of the nonlinear EEG features that were investigated in the present study and thus aid in the interpretation of the results.

Third, assume dimensionality. One of the key principles of the RDoC approach is that behaviourally and neurobiologically defined variables have no clear clinical cut-off but range from normal to abnormal in a continuous manner. This does not mean that classification is impossible but these should be based on naturally occurring and empirically detected cut-offs as opposed to an arbitrary number of symptoms. Future studies collecting both behavioural and neurobiological data (e.g. nonlinear EEG measures) should investigate the degree to which both measures scale-up (e.g. Bosl et al., 2018).

The use of enormous data-sets of neuroimaging data in order to help the diagnosis of mental disorders is putting the cart before the horse. While the methodology in and of itself is useful, it is too early to harness its full potential. An approach in which we first focus our efforts on the basis, the symptoms itself and attempt to discern the neurobiological basis of these symptoms, appreciating its dimensional character, is more likely to yield productive results. Only then will we be able to reliably predict behavioural deficits using neuroimaging at an early age. For us, as a research community, to arrive at such a stage, it will take years of rigorous experimentation, documentation and many readjustments along the path in which we put the symptoms first instead of the clinical category.

Conclusion

The present study investigated the potential of nonlinear EEG features as an early biomarker of ASD. A classic type of machine learning algorithms, namely support vector machines, was implemented in the classification of infants as either typically developing or in need of further monitoring based on their EEG features. Analysis revealed that these machine learning algorithms were not able to identify the EEG features of infants with ASD with satisfactory sensitivity or specificity. No significant differences were found between the group with a positive screening for ASD and the group with a negative screening for ASD on the EEG features at both 10 and 14 months old. While longitudinal designs, more advanced analysis methods and more data might solve some issues, the main perpetrator in halting our understanding of the neurobiology of ASD is the diagnosis itself. We therefore outline three ways in which future studies can inform us on the neurobiology of ASD by adhering to the RDoC research framework. First, by focussing on the psychological constructs themselves instead of on a diagnosis. Second, by integrating the biological and behavioural measurements in a meaningful context (i.e. more naturalistic experiments). Lastly, by assuming dimensionality of both the behavioural as the neurobiological variables. We believe that by focussing on the basis, the symptoms itself and its neurobiological underpinnings, the field will make large strides in its understanding of developmental disorders and their early prediction.

References

Abdi, H. (2010). Holm's sequential Bonferroni procedure. *Encyclopedia of research design*, *1*(8), 1-8.

Agrawal, S., Rao, S. C., Bulsara, M. K., & Patole, S. K. (2018). Prevalence of autism spectrum disorder in preterm infants: a meta-analysis. *Pediatrics*, *142*(3).

Ahearn, W. H., Castine, T., Nault, K., & Green, G. (2001). An assessment of food acceptance in children with autism or pervasive developmental disorder-not otherwise specified. *Journal of Autism and Developmental Disorders*, *31*(5), 505-511.

Allison, C., Auyeung, B., & Baron-Cohen, S. (2012). Toward brief "red flags" for autism screening: the short autism spectrum quotient and the short quantitative checklist in 1,000 cases and 3,000 controls. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(2), 202-212. e207.

Allison, C., Baron-Cohen, S., Wheelwright, S., Charman, T., Richler, J., Pasco, G., & Brayne, C. (2008). The Q-CHAT (Quantitative CHecklist for Autism in Toddlers): a normally distributed quantitative measure of autistic traits at 18–24 months of age: preliminary report. *Journal of Autism and Developmental Disorders*, *38*(8), 1414-1425.

American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.). American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). American Psychiatric Association. https://doi.org/10.1176/appi.books.9780890425596

Babbitt, R. L., Hoch, T. A., Coe, D. A., Cataldo, M. F., Kelly, K. J., Stackhouse, C., & Perman, J. A. (1994). Behavioral assessment and treatment of pediatric feeding disorders. *Journal of developmental and Behavioral Pediatrics*.

Ball, G., Boardman, J. P., Aljabar, P., Pandit, A., Arichi, T., Merchant, N., Rueckert, D., Edwards, A. D., & Counsell, S. J. (2013). The influence of preterm birth on the developing thalamocortical connectome. *Cortex*, *49*(6), 1711-1721.

Baron-Cohen, S., Allen, J., & Gillberg, C. (1992). Can autism be detected at 18 months?: The needle, the haystack, and the CHAT. *The British Journal of Psychiatry*, *161*(6), 839-843.

Baron-Cohen, S., Cox, A., Baird, G., Swettenham, J., Nightingale, N., Morgan, K., Drew, A., & Charman, T. (1996). Psychological markers in the detection of autism in infancy in a large population. *The British Journal of Psychiatry*, *168*(2), 158-163.

Barttfeld, P., Amoruso, L., Ais, J., Cukier, S., Bavassi, L., Tomio, A., Manes, F., Ibanez, A., & Sigman, M. (2013). Organization of brain networks governed by long-range connections index autistic traits in the general population. *Journal of neurodevelopmental disorders*, *5*(1), 16. https://doi.org/10.1186/1866-1955-5-16

Bieleninik, L., Posserud, M. B., Geretsegger, M., Thompson, G., Elefant, C., & Gold, C. (2017). Tracing the temporal stability of autism spectrum diagnosis and severity as measured by the Autism Diagnostic Observation Schedule: A systematic review and meta-analysis. *PLOS ONE*, *12*(9), e0183160. https://doi.org/10.1371/journal.pone.0183160

Bokobza, C., Van Steenwinckel, J., Mani, S., Mezger, V., Fleiss, B., & Gressens, P. (2019). Neuroinflammation in preterm babies and autism spectrum disorders. *Pediatric Research*, *85*(2), 155-165.

Bondy, A., & Frost, L. (2001). The Picture Exchange Communication System. *Behavior modification*, *25*(5), 725-744. https://doi.org/10.1177/0145445501255004

Bosl, W., Tierney, A., Tager-Flusberg, H., & Nelson, C. (2011). EEG complexity as a biomarker for autism spectrum disorder risk. *BMC Medicine*, *9*(18). https://doi.org/10.1186/1741-7015-9-18

Bosl, W. J., Tager-Flusberg, H., & Nelson, C. A. (2018). EEG Analytics for Early Detection of Autism Spectrum Disorder: A data-driven approach. *Scientific Reports 8*, 6828. https://doi.org/10.1038/s41598-018-24318-x

Bruni, T. P. (2014). Test review: Social responsiveness scale–Second edition (SRS-2). *Journal of Psychoeducational Assessment*, *32*(4), 365-369.

Bryce, R. M., & Sprague, K. B. (2012). Revisiting detrended fluctuation analysis. *Scientific Reports*, *2*(1), 315. https://doi.org/10.1038/srep00315

Butler, S., & Lord, C. (2013). Rimland Diagnostic Form for Behavior-Disturbed Children (E-2). In F.R. Volkmar (Ed.), *Encyclopedia of Autism Spectrum Disorders*. Springer. https://doi.org/10.1007/978-1-4419-1698-3_914

Camarata, S. (2014). Early identification and early intervention in autism spectrum disorders: Accurate and effective? *International Journal of Speech-Language Pathology*, *16*(1), 1-10. https://doi.org/10.3109/17549507.2013.858773

Cateni, S., Vannucci, M., Vannocci, M., & Valentina, G. (2013). Variable Selection and Feature Extraction Through Artificial Intelligence Techniques. In *Multivariate Analysis in Management, Engineering and the Science* (pp. 103-118).

Chiarotti, F., & Venerosi, A. (2020). Epidemiology of Autism Spectrum Disorders: A Review of Worldwide Prevalence Estimates Since 2014. *Brain Sciences*, *10*(5), 274. https://doi.org/10.3390/brainsci10050274

Clarkson, T., Kang, E., Capriola-Hall, N., Lerner, M. D., Jarcho, J., & Prinstein, M. J. (2020). Meta-analysis of the RDoC social processing domain across units of analysis in children and adolescents. *Journal of Clinical Child & Adolescent Psychology*, *49*(3), 297-321.

Cohen, M. X. (2017). Where does EEG come from and what does it mean? *Trends in neurosciences*, *40*(4), 208-218.

Constantino, J. N., & Gruber, C. P. (2012). *Social responsiveness scale: SRS-2*. Western Psychological Services.

Cornew, L., Roberts, T. P. L., Blaskey, L., & Edgar, J. C. (2012). Resting-State Oscillatory Activity in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *42*(9), 1884-1894. https://doi.org/10.1007/s10803-011-1431-6

Dawson, G., Klinger, L. G., Panagiotides, H., Lewy, A., & Castelloe, P. (1995). Subgroups of autistic children based on social behavior display distinct patterns of brain activity. *Journal of Abnormal Child Psychology*, *23*(5), 569-583. https://doi.org/10.1007/BF01447662

De Bildt, A., De Jonge, M., & Greaves-Lord, K. (2013). ADOS-2 autism diagnostic observation schedule (Dutch version). *Amsterdam, the Netherlands: Hogrefe.*

Doernberg, E., & Hollander, E. (2016). Neurodevelopmental disorders (asd and adhd): dsm-5, icd-10, and icd-11. *CNS spectrums*, *21*(4), 295-299.

Elsabbagh, M. (2020). Linking risk factors and outcomes in autism spectrum disorder: is there evidence for resilience? *BMJ*, *368*. https://doi.org/10.1136/bmj.l6880

Elsabbagh, M., Divan, G., Koh, Y. J., Kim, Y. S., Kauchali, S., Marcín, C., Montiel-Nava, C., Patel, V., Paula, C. S., Wang, S., Yasamy, M. T., & Fombonne, E. (2012). Global Prevalence of Autism and Other Pervasive Developmental Disorders. *Autism Research*, *5*(3), 160-179. https://doi.org/10.1002/aur.239

Goldenberg, R. L., Culhane, J. F., Iams, J. D., & Romero, R. (2008). Epidemiology and causes of preterm birth. *The lancet*, *371*(9606), 75-84.

Golland, P., & Fischl, B. (2003). Permutation tests for classification: towards statistical significance in image-based studies. Biennial international conference on information processing in medical imaging,

Gotham, K., Pickles, A., & Lord, C. (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*(5), 693-705. https://doi.org/10.1007/s10803-008-0674-3

Gregory, L. M., Burton, V. J., & Shapiro, B. K. (2015). Chapter 3 - Developmental Disabilities and Metabolic Disorders. In M. J. Zigmond, L. P. Rowland, & J. T. Coyle (Eds.), *Neurobiology of Brain Disorders* (pp. 18-41). Academic Press. https://doi.org/10.1016/B978-0-12-398270-4.00003-3

Gurau, O., Bosl, W. J., & Newton, C. R. (2017). How useful is electroencephalography in the diagnosis of autism spectrum disorders and the delineation of subtypes: a systematic review. *Frontiers in Psychiatry*, *8*, 121.

Harrison, L. A., Kats, A., Kilroy, E., Butera, C., Jayashankar, A., Keles, U., & Aziz-Zadeh, L. (2021). Motor and sensory features successfully decode autism spectrum disorder and combine with the original RDoC framework to boost diagnostic classification. *Scientific Reports*, *11*(1), 1-16.

Hennessey, T., Andari, E., & Rainnie, D. G. (2018). RDoC-based categorization of amygdala functions and its implications in autism. *Neuroscience & Biobehavioral Reviews*, *90*, 115-129.

Hertz-Picciotto, I., Croen, L. A., Hansen, R., Jones, C. R., van de Water, J., & Pessah, I. (2006). The CHARGE Study: An Epidemiologic Investigation of Genetic and Environmental Factors Contributing to Autism. *Environmental Health Perspectives*, *114*(7). https://doi.org/10.1289/ehp.8483

Hirata, Y., & Aihara, K. (2011). STATISTICAL TESTS FOR SERIAL DEPENDENCE AND LAMINARITY ON RECURRENCE PLOTS. *International Journal of Bifurcation and Chaos*, *21*(04), 1077-1084. https://doi.org/10.1142/s0218127411028908

Horwitz, B. (2003). The elusive concept of brain connectivity. *NeuroImage*, *19*(2), 466-470. https://doi.org/10.1016/S1053-8119(03)00112-5

Hyde, K. K., Novack, M. N., LaHaye, N., Parlett-Pelleriti, C., Anden, R., Dixon, D. R., & Linstead, E. (2019). Applications of Supervised Machine Learning in Autism Spectrum Disorder Research: a Review. *Review Journal of Autism and Developmental Disorders*, *6*(2), 128-146. https://doi.org/10.1007/s40489-019-00158-x

Hyman, S. L., Levy, S. E., & Myers, S. M. (2020). Identification, Evaluation, and Management of Children With Autism Spectrum Disorder. *Pediatrics*, *145*(1), e20193447. https://doi.org/10.1542/peds.2019-3447

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*.

Keen, D. V. (2008). Childhood autism, feeding problems and failure to thrive in early infancy. *European child & adolescent psychiatry*, *17*(4), 209-216.

Khalid, S., Khalil, T., & Nasreen, S. (2014, 27-29 Aug. 2014). A survey of feature selection and feature extraction techniques in machine learning. 2014 Science and Information Conference,

Koegel, R. L., Koegel, L. K., & McNerney, E. K. (2001). Pivotal areas in intervention for autism. *Journal of clinical child psychology*, *30*(1), 19-32. https://doi.org/10.1207/s15374424jccp3001_4

Kwaliteitscentrum voor Diagnostiek vzw (2018). Classificerend Diagnostisch Protocol Autismespectrumstoornis bij kinderen en jongeren. Retrieved (21/05/2022) from http://www.kwaliteitscentrumdiagnostiek.be/publicatiesjeugdhulp

Ladha, L., & Deepa, T. (2011). Feature selection methods and algorithms. *International journal on computer science and engineering*, *3*(5), 1787-1797.

Lazarev, V. V., Pontes, A., & deAzevedo, L. C. (2009). EEG photic driving: Right-hemisphere reactivity deficit in childhood autism. A pilot study. *International Journal of Psychophysiology*, *71*(2), 177-183. https://doi.org/https://doi.org/10.1016/j.ijpsycho.2008.08.008

Ledford, J. R., & Gast, D. L. (2006). Feeding problems in children with autism spectrum disorders: A review. *Focus on Autism and Other Developmental Disabilities*, *21*(3), 153-166.

Leviton, A., Joseph, R. M., Allred, E. N., Fichorova, R. N., O'Shea, T. M., Kuban, K. K., & Dammann, O. (2018). The risk of neurodevelopmental disorders at age 10 years associated with blood concentrations of interleukins 4 and 10 during the first postnatal month of children born extremely preterm. *Cytokine*, *110*, 181-188.

Li, Z., Zhang, L., Zhang, F., Gu, R., Peng, W., & Hu, L. (2020). Demystifying signal processing techniques to extract resting- state EEG features for psychologists. *Brain Science Advances*, *6*(3), 189-209. https://doi.org/10.26599/bsa.2020.9050019

Linting, M., Meulman, J. J., Groenen, P. J. F., & van der Koojj, A. J. (2007). Nonlinear principal components analysis: introduction and application. *Psychological methods*, *12*(3), 336.

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism Diagnostic Observation Schedule, Second Edition*. Western Psychological Services.

Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology*, *55*(1), 3-9. https://doi.org/10.1037/0022-006X.55.1.3

Luyster, R., Gotham, K., Guthrie, W., Coffing, M., Petrak, R., Pierce, K., Bishop, S., Esler, A., Hus, V., & Oti, R. (2009). The Autism Diagnostic Observation Schedule—Toddler Module: A new module of a standardized diagnostic measure for autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*(9), 1305-1320.

Magiati, I., & Howlin, P. (2001). Monitoring the Progress of Preschool Children with Autism Enrolled in Early Intervention ProgrammesProblems in Cognitive Assessment. *Autism*, *5*, 399-406. https://doi.org/10.1177/1362361301005004005

Marwan, N. (2011). HOW TO AVOID POTENTIAL PITFALLS IN RECURRENCE PLOT BASED DATA ANALYSIS. *International Journal of Bifurcation and Chaos*, *21*(04), 1003-1017. https://doi.org/10.1142/s0218127411029008

Marwan, N. (2012). *Recurrence Plots and Cross Recurrence Plots*. Retrieved 12/04/2021 from http://www.recurrence-plot.tk

Marwan, N., Carmen Romano, M., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports*, *438*(5), 237-329. https://doi.org/https://doi.org/10.1016/j.physrep.2006.11.001

Marwan, N., Wessel, N., Meyerfeldt, U., Schirdewan, A., & Kurths, J. (2002). Recurrence-plot-based measures of complexity and their application to heart-rate-variability data. *Physical Review E*, *66*(2), 026702. https://doi.org/10.1103/PhysRevE.66.026702

Matson, J. L., Belva, B. C., Horovitz, M., Kozlowski, A. M., & Bamburg, J. W. (2012). Comparing Symptoms of Autism Spectrum Disorders in a Developmentally Disabled Adult Population Using the Current DSM-IV-TR Diagnostic Criteria and the Proposed DSM-5 Diagnostic Criteria. *Journal*

of Developmental and Physical Disabilities, *24*, 403-414. https://doi.org/10.1007/s10882-012-9278-0

McCrimmon, A., & Rostad, K. (2013). Test Review: Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) Manual (Part II): Toddler Module. *Journal of Psychoeducational Assessment*, *32*(1), 88-92. https://doi.org/10.1177/0734282913490916

McPartland, J. C., Reichow, B., & Volkmar, F. R. (2012). Sensitivity and Specificity of Proposed DSM-5 Diagnostic Criteria for Autism Spectrum Disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(4). https://doi.org/10.1016/j.jaac.2012.01.007

McVey, A. J. (2019). The neurobiological presentation of anxiety in autism spectrum disorder: A systematic review. *Autism Research*, *12*(3), 346-369.

Narzisi, A., Costanza, C., Umberto, B., & Filippo, M. (2014). Non-Pharmacological Treatments in Autism Spectrum Disorders: An Overview on Early Interventions for Pre-Schoolers. *Current Clinical Pharmacology*, *9*(1), 17-26. https://doi.org/10.2174/15748847113086660071

National Institute of Mental Health. *Research Domain Criteria (RDoC)*. NIMH website. Retrieved 02/05/2022 from https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc

O'Reilly, C., Lewis, J. D., & Elsabbagh, M. (2017). Is functional brain connectivity atypical in autism? A systematic review of EEG and MEG studies. *PLOS ONE*, *12*(5), e0175870. https://doi.org/10.1371/journal.pone.0175870

Orekhova, E. V., Elsabbagh, M., Jones, E. J. H., Dawson, G., Charman, T., Johnson, M. H., & The BASIS Team. (2014). EEG hyper-connectivity in high-risk infants is associated with later autism. *Journal of neurodevelopmental disorders*, *6*(40). https://doi.org/10.1186/1866-1955-6-40

Ozonoff, S., Young, G. S., Carter, A., Messinger, D., Yirmiya, N., Zwaigenbaum, L., Bryson, S., Carver, L. J., Constantino, J. N., Dobkins, K., Hutman, T., Iverson, J. M., Landa, R., Rogers, S. J., Sigman, M., & Stone, W. L. (2011). Recurrence Risk for Autism Spectrum Disorders: A Baby Siblings Research Consortium Study. *Pediatrics*, *128*(3), e488-e495. https://doi.org/10.1542/peds.2010-2825

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, 2825-2830.

Prizant, B. M., Wetherby, A., Rubin, E., & Laurent, A. C. (2003). The SCERTS Model: A Transactional, Family-Centered Approach to Enhancing Communication and Socio emotional Abilities of Children With Autism Spectrum Disorder. *Infants and Young Children*, *16*(4), 296-316.

Raichle, M. E., & Snyder, A. Z. (2007). A default mode of brain function: A brief history of an evolving idea. *NeuroImage*, *37*(4), 1083-1090. https://doi.org/10.1016/j.neuroimage.2007.02.041

Randall, M., Egberts, K. J., Samtani, A., Scholten, R. J. P. M., Hooft, L., Livingstone, N., Sterling-Levis, K., Woolfenden, S. & Williams, K. (2018). *Diagnostic tests for autism spectrum disorder (ASD) in*

preschool children Cochrane Database of Systematic Reviews. https://doi.org/10.1002/14651858.CD009044.pub2

Reichow, B. (2012). Overview of Meta-Analyses on Early Intensive Behavioral Intervention for Young Children with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *42*(4), 512-520. https://doi.org/10.1007/s10803-011-1218-9

Renty, J. O., & Roeyers, H. (2006). Quality of life in high-functioning adults with autism spectrum disorder: The predictive value of disability and support characteristics. *Autism*, *10*(5), 511-524. https://doi.org/10.1177/1362361306066604

Richman, J. S., & Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology*, *278*(6), H2039-H2049. https://doi.org/10.1152/ajpheart.2000.278.6.H2039

Roberts, T. P. L., Khan, S. Y., Rey, M., Monroe, J. F., Cannon, K., Blaskey, L., Woldoff, S., Qasmieh, S., Gandal, M., Schmidt, G. L., Zarnow, D. M., Levy, S. E., & Edgar, J. C. (2010). MEG detection of delayed auditory evoked responses in autism spectrum disorders: towards an imaging biomarker for autism. *Autism Research*, *3*(1), 8-18. https://doi.org/https://doi.org/10.1002/aur.111

Robins, D. L., Fein, D., & Barton M. L. (1999). *The modified checklist for autism in toddlers (M-CHAT)*. Self-published.

Rodgers, M., Simmonds, M., Marshall, D., Hodgson, R., Stewart, L. A., Rai, D., Wright, K., Ben-Itzchak, E., Eikeseth, S., & Eldevik, S. (2021). Intensive behavioural interventions based on applied behaviour analysis for young children with autism: An international collaborative individual participant data meta-analysis. *Autism*, 1362361320985680.

Rohilla, A. (2018, November 2). *A brief introduction to Support Vector machine*. Towards Data Science. Retrieved 09/04/2021 from https://towardsdatascience.com/a-brief-introduction-to-support-vector-machine-adf0f103a80f#:~:text=In%20SVM%2C%20to%20avoid%20overfitting,overfit%20on%20our%20training%20sample

Schinkel, S., Marwan, N., & Kurths, J. (2009). Brain signal analysis based on recurrences. *Journal of Physiology-Paris*, *103*(6), 315-323. https://doi.org/https://doi.org/10.1016/j.jphysparis.2009.05.007

Tay, T. L., Savage, J. C., Hui, C. W., Bisht, K., & Tremblay, M. È. (2017). Microglia across the lifespan: from origin to function in brain development, plasticity and cognition. *The Journal of physiology*, *595*(6), 1929-1945.

Tick, B., Bolton, P., Happé, F., Rutter, M., & Rijsdijk, F. (2015). Heritability of autism spectrum disorders: a meta-analysis of twin studies. *J Child Psychol Psychiatry*, *57*(5), 585-595. https://doi.org/10.1111/jcpp.12499

Tierney, A. L., Gabard-Durnam, L., Vogel-Farley, V., Tager-Flusberg, H., & Nelson, C. A. (2012). Developmental Trajectories of Resting EEG Power: An Endophenotype of Autism Spectrum Disorder. *PLOS ONE*, *7*(6). https://doi.org/10.1371/journal.pone.0039127

Tsai, L. Y. (2014). Impact of DSM-5 on epidemiology of Autism Spectrum Disorder. *Research in Autism Spectrum Disorders*, *8*(11). https://doi.org/10.1016/j.rasd.2014.07.016

Tschida, J. E., & Yerys, B. E. (2021). A systematic review of the positive valence system in autism spectrum disorder. *Neuropsychology Review*, *31*(1), 58-88.

van Heijst, B. F. C., & Geurts, H. M. (2015). Quality of life in autism across the lifespan: A meta-analysis [Article]. *Autism*, *19*(2), 158-167. https://doi.org/10.1177/1362361313517053

Vismara, L. A., & Rogers, S. J. (2010). Behavioral Treatments in Autism Spectrum Disorder: What Do We Know? *Annual Review of Clinical Psychology 6*, 447-468. https://doi.org/10.1146/annurev.clinpsy.121208.131151

Vissers, M. E., Cohen, M. X., & Geurts, H. M. (2012). Brain connectivity and high functioning autism: A promising path of research that needs refined models, methodological convergence, and stronger behavioral links. *Neuroscience & Biobehavioral Reviews*, *36*(1), 604-625. https://doi.org/10.1016/j.neubiorev.2011.09.003

Volkmar, F. R., & McPartland, J. C. (2014). From Kanner to DSM-5: Autism as an Evolving Diagnostic Concept. *Annual Review of Clinical Psychology*, *10*(1), 193-212. https://doi.org/10.1146/annurev-clinpsy-032813-153710

Wang, J., Barstein, J., Ethridge, L. E., Mosconi, M. W., Takarae, Y., & Sweeney, J. A. (2013). Resting state EEG abnormalities in autism spectrum disorders. *Journal of neurodevelopmental disorders*, *5*(1), 24-24. https://doi.org/10.1186/1866-1955-5-24

Warren, Z., McPheeters, M. L., Sathe, N., Foss-Feig, J. H., Glasser, A., & Veenstra-VanderWeele, J. (2011). A Systematic Review of Early Intensive Intervention for Autism Spectrum Disorders. *Pediatrics*, *127*(5), e1303. https://doi.org/10.1542/peds.2011-0426

Weitlauf, A. S., Vehorn, A. C., Stone, W. L., Fein, D., & Warren, Z. E. (2015). Using the M-CHAT-R/F to identify developmental concerns in a high-risk 18-month-old sibling sample. *Journal of developmental and behavioral pediatrics: JDBP*, *36*(7), 497.

Wetherby, A., & Woods, J. (2006). Early Social Interaction Project for Children With Autism Spectrum Disorders Beginning in the Second Year of Life: A Preliminary Study. *Topics in Early Childhood Special Education 26*, 67-82. https://doi.org/10.1177/02711214060260020201

Williams, J. G., Higgins, J. P. T., & Brayne, C. E. G. (2006). Systematic review of prevalence studies of autism spectrum disorders. *Archives of Disease in Childhood*, *91*, 8-15. https://doi.org/10.1136/adc.2004.062083

Wong, E. H. F., Fox, J. C., Ng, M. Y. M., & Lee, C. (2011). 13 - Toward Personalized Medicine in the Neuropsychiatric Field. In P. C. Guest & S. Bahn (Eds.), *International Review of Neurobiology* (Vol. 101, pp. 329-349). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-387718-5.00013-4

Worley, J. A., & Matson, J. L. (2012). Comparing symptoms of autism spectrum disorders using the current DSM-IV-TR diagnostic criteria and the proposed DSM-V diagnostic criteria. *Res. Autism Spectr. Disord*, *6*, 965-970. https://doi.org/10.1016/j.rasd.2011.12.012

Yeh, C.C., Chen, C.Y., & Wang, P.H. (2017). Infection and preterm birth. In (Vol. 80, pp. 530-531): LWW.

Zbilut, J. P., & Webber, C. L. (1992). Embeddings and delays as derived from quantification of recurrence plots. *Physics Letters A*, *171*(3), 199-203. https://doi.org/https://doi.org/10.1016/0375-9601(92)90426-M

Zeidan, J., Fombonne, E., Scorah, J., Ibrahim, A., Durkin, M. S., Saxena, S., Yusuf, A., Shih, A., & Elsabbagh, M. (2022). Global prevalence of autism: A systematic review update. *Autism Research*.